



ASTES

Advances in Science, Technology & Engineering Systems Journal

VOLUME 7-ISSUE 2 | MAR-APR 2022

www.astesj.com

ISSN: 2415-6698

EDITORIAL BOARD

Editor-in-Chief

Prof. Passerini Kazmerski
University of Chicago, USA

Editorial Board Members

Dr. Jiantao Shi

Nanjing Research Institute
of Electronic Technology,
China

Dr. Tariq Kamal

University of Nottingham, UK
Sakarya University, Turkey

Dr. Mohmaed Abdel Fattah

Ashabrawy
Prince Sattam bin Abdulaziz University,
Saudi Arabia

Dr. Nguyen Tung Linh

Electric Power University,
Vietnam

**Prof. Majida Ali Abed
Meshari**

Tikrit University Campus,
Iraq

Mr. Muhammad Tanveer Riaz

School of Electrical Engineering,
Chongqing University, P.R. China

**Mohamed Mohamed
Abdel-Daim**

Suez Canal University,
Egypt

Dr. Omeje Maxwell

Covenant University, Nigeria

Dr. Hung-Wei Wu

Kun Shan University, Taiwan

Dr. Heba Afify

MTI university, Cairo, Egypt

Mr. Randhir Kumar

National University of
Technology Raipur, India

Dr. Ahmet Kayabasi Karamanoglu

Mehmetbey University, Turkey

Dr. Daniele Mestriner

University of Genoa, Italy

Dr. Hongbo Du

Prairie View A&M
University, USA

Mr. Aamir Nawaz

Gomal University, Pakistan

Dr. Abhishek Shukla

R.D. Engineering College, India

Mr. Manu Mitra

University of Bridgeport, USA

Regional Editors

Dr. Maryam Asghari

Shahid Ashrafi Esfahani,
Iran

Mr. Abdullah El-Bayoumi

Cairo University, Egypt

Dr. Sabry Ali Abdallah El-Naggar

Tanta University, Egypt

Dr. Ebubekir Altuntas

Gaziosmanpasa University,
Turkey

Dr. Qichun Zhang

University of Bradford,
United Kingdom

Dr. Walid Wafik Mohamed Badawy

National Organization for Drug Control
and Research, Egypt

Dr. Gomathi Periasamy

Mekelle University, Ethiopia

Dr. Shakir Ali

Aligarh Muslim University,
India

Dr. Ayham Hassan Abazid

Jordan
University of Science and Technology,
Jordan

Editorial

We are delighted to present this issue featuring 19 accepted research articles spanning a wide range of engineering and technology topics.

The issue begins with Adekunle and Adewale's paper on implementing and evaluating wireless sensor networks for automated hydroponic systems [1]. Through simulations, they found cluster-based networks provided lower latency and energy consumption compared to multi-hop networks as the system scaled up. These findings can guide future smart agriculture systems.

Shifting focus to vehicular networks, Laouiti et al. put forth an enhanced cross-layer mechanism for real-time high-efficiency HEVC video streaming over VANETs [2]. By adaptive video frame prioritization based on network conditions, their approach achieved improved video quality and lower latency compared to standard methods. This has important implications for future connected vehicles.

In the field of network security, Aissani et al. present a taxonomy of techniques for securing routing protocols in mobile ad-hoc networks [3]. Analyzing cryptography-based and trust-based solutions, they highlight the need for hybrid methods leveraging both encryption and reputation systems to ensure node honesty. Their work provides a useful framework for MANET security.

Renewable energy systems are explored in El Bakkali et al.'s paper on PV-battery hybrid grid integration [4]. They developed smart power management strategies aimed at ensuring reliable supply while preventing battery overcharging/discharging. Simulations verified their approach can reduce customer energy costs. This provides a model for sustainable hybrid grids.

Shifting focus to project management, Laaz et al. study IT project models during the era of digital transformation [5]. Identifying constraints posed by new technologies, they propose a hybrid framework blending aspects of traditional waterfall and agile methodologies. Practical insights are offered for adapting to the digital age.

In the power electronics domain, Rathore et al. analyze the stability of DC microgrids with constant power loads [6]. Through theoretical analysis and simulations, they derive power level guidelines and current controller bandwidths that maintain stability without requiring additional compensators. This economical approach facilitates integration of new load types.

Probabilistic logic modeling is the theme of Schulte et al.'s work on constructing and utilizing symmetries in dynamic relational domains [7]. They introduce novel techniques to detect object behavioral symmetries over time, preventing unnecessary grounding and enabling more efficient lifted inference. This advances graphical modeling capabilities.

The study by Kissane et al. investigates blended learning for tertiary mathematics, replacing a face-to-face tutorial with online activities [8]. Their quasi-experimental results showed improved performance for entering students and similar outcomes for more advanced students. This supports incorporating e-learning alongside traditional methods.

Shifting focus to renewable assessment, Mudau et al. present solar radiation modeling and estimation using weather data in South Africa's Vhembe District [9]. Applying data analytics and mapping approaches provides insight into local solar energy potential. Their work aids planning of renewable generation systems.

Computer vision for healthcare is spotlighted in Ito et al.'s paper on using generative adversarial networks to evaluate hand hygiene [10]. Using fluorescent imaging of washed hands, their deep learning model achieved high accuracy in detecting cleaning thoroughness. This demonstrates an innovative application of AI to improve training and practices.

Automatic image captioning is explored by Chaudhari et al. through a unified visual saliency framework [11]. Integrating convolutional and recurrent neural networks with attention modeling, their approach provides state-of-the-art performance on both general and medical image datasets. Advancing multi-modal understanding has broad impacts.

In the field of digital forensics, Alotaibi presents techniques for recovering WhatsApp artifacts on Android devices without root access [12]. Their forensic analysis identified valuable messaging evidence extractable from internal storage. Practical tools to access encrypted app data aid law enforcement investigations.

Taking an interdisciplinary perspective, Lebyodkin articulates a condensed matter physics approach to modeling fracture in solids [13]. Adopting rheological material models and micro-damage based failure criteria, their framework has potential for predicting component durability under varied loading conditions. This synergistic view may unlock new insights.

Decision support through interpretable AI is the focus of Imamori et al.'s work on explaining mastitis detection models for dairy cows [14]. Applying inductive logic programming, they extracted symbolic rules characterizing influential patterns in sensor measurements. Increased model transparency and trust facilitates practical adoption.

Shifting to public health technology, Al-Ruithe et al. provide an updated discussion of contact prevention mobile apps amidst the continued COVID-19 pandemic [15]. They argue such tools remain beneficial for mitigating transmission and propose features enhancing user awareness. Ongoing innovation is critical as viruses evolve.

Data security is revisited in El Bilali et al.'s examination of privacy for forward private searchable encryption [16]. They introduce new leakage-abuse attacks recovering search queries from access patterns and mitigate via obfuscation techniques. Advancing cryptographic protections is crucial as data volumes grow.

Assessing education outcomes, Alelyani et al. analyze digital readiness among Saudi university students [17]. Their survey highlights competence gaps, with lower skills in safety and content creation. Recommendations are presented for improving curriculum and better cultivating digital citizenship.

Designing scalable blockchain frameworks for healthcare data sharing is explored by Wazid et al. [18]. They propose federated hospital-city-state networks with access controls facilitating efficient cross-organization EHR queries. Reliable and timely data availability can improve care quality.

Rounding out the issue, Norman examines application of strategic design for organizational transformation [19]. A contextual framework is developed identifying links between operational and management activities where interventions and measurement can enable innovation capabilities. This provides useful guidance for companies pursuing change.

In summary, the excellent breadth of topics covered in this special issue exemplify high-caliber research advancing the frontiers of science, engineering and technology. We hope these

contributions will stimulate further interdisciplinary studies and translations of ideas into impactful innovations. We thank the authors for entrusting their work to our journal and our reviewers for upholding rigorous standards.

References:

- [1] M. Musa, A. Mabu, F. Modu, A. Adam, F. Aliyu, "Automated Hydroponic System using Wireless Sensor Networks," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 1–17, 2022, doi:10.25046/aj070201.
- [2] M. Hassan, A. Badri, A. Sahel, B. Kochairi, N. Baghdad, "Enhanced Dynamic Cross Layer Mechanism for real time HEVC Streaming over Vehicular Ad-hoc Networks (VANETs)," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 18–24, 2022, doi:10.25046/aj070202.
- [3] K. Zaid, D. Ouafaa, "Taxonomy of Security Techniques for Routing Protocols in Mobile Ad-hoc Networks," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 25–31, 2022, doi:10.25046/aj070203.
- [4] O. Mohammed, L.M. Tahar, L. Nora, "Power Management and Control of a Grid-Connected PV/Battery Hybrid Renewable Energy System," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 32–52, 2022, doi:10.25046/aj070204.
- [5] R. Hassani, Y.E.B. El Idrissi, "IT Project Management Models in an Era of Digital Transformation: A Study by Practice," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 53–62, 2022, doi:10.25046/aj070205.
- [6] S. Ansari, K. Iqbal, "Stability Analysis of a DC Microgrid with Constant Power Load," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 63–72, 2022, doi:10.25046/aj070206.
- [7] N. Finke, R. Möller, "On the Construction of Symmetries and Retaining Lifted Representations in Dynamic Probabilistic Relational Models," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 73–93, 2022, doi:10.25046/aj070207.
- [8] M.R. Freislich, A. Bowen-James, "Online Support for Tertiary Mathematics Students in a Blended Learning Environment," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 94–102, 2022, doi:10.25046/aj070208.
- [9] C. Matasane, M.T. Kahn, "Solar Energy Assessment, Estimation, and Modelling using Climate Data and Local Environmental Conditions," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 103–111, 2022, doi:10.25046/aj070209.
- [10] F. Kinoshita, K. Nagano, G. Cui, M. Yoshii, H. Touyama, "A Study on Novel Hand Hygiene Evaluation System using pix2pix," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 112–118, 2022, doi:10.25046/aj070210.
- [11] S.S.P.R. Amma, S.M. Idicula, "A Unified Visual Saliency Model for Automatic Image Description Generation for General and Medical Images," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 119–126, 2022, doi:10.25046/aj070211.
- [12] M. Shadeed, L.A. Arram, M. Owda, "Forensic Analysis of 'WhatsApp' Artifacts in Android without Root," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 127–132, 2022, doi:10.25046/aj070212.
- [13] M. Petrov, "An Interdisciplinary Approach to Fracture of Solids from the Standpoint of Condensed Matter Physics," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 133–142, 2022, doi:10.25046/aj070213.
- [14] H. Motohashi, H. Ohwada, "Interpretable Rules Using Inductive Logic Programming Explaining Machine Learning Models: Case Study of Subclinical Mastitis Detection for Dairy Cows," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 143–148, 2022, doi:10.25046/aj070214.

- [15] E.M. Mimo, T. McDaniel, J.B. Ruvunangiza, "COVIDFREE App: The User-Enabling Contact Prevention Application: A Review," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 149–155, 2022, doi:10.25046/aj070215.
- [16] K. Salmani, K. Barker, "Leakage-abuse Attacks Against Forward Private Searchable Symmetric Encryption," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 156–170, 2022, doi:10.25046/aj070216.
- [17] I. Abousaber, "Digital Competencies of Saudi University Graduates Towards Digital Society: The Case of The University of Tabuk," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 171–178, 2022, doi:10.25046/aj070217.
- [18] A. Thamrin, H. Xu, R. Ming, "Cloud-Based Hierarchical Consortium Blockchain Networks for Timely Publication and Efficient Retrieval of Electronic Health Records," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 179–190, 2022, doi:10.25046/aj070218.
- [19] L. Whelan, L. Kiernan, K. Morrissey, N. Deloughry, "Towards a Framework for Organizational Transformation through Strategic Design Implementation," *Advances in Science, Technology and Engineering Systems Journal*, **7(2)**, 191–197, 2022, doi:10.25046/aj070219.

Editor-in-chief

Prof. Passerini Kazmersk

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Volume 7 Issue 2

March-April 2022

CONTENTS

<i>Automated Hydroponic System using Wireless Sensor Networks</i> Mahdi Musa, Audu Mabu, Falmata Modu, Adam Adam, Farouq Aliyu	01
<i>Enhanced Dynamic Cross Layer Mechanism for real time HEVC Streaming over Vehicular Ad-hoc Networks (VANETs)</i> Marzouk Hassan, Abdelmajid Badri, Aicha Sahel, Belbachir Kochairi, Nacer Baghdad	18
<i>Taxonomy of Security Techniques for Routing Protocols in Mobile Ad-hoc Networks</i> Kartit Zaid, Diouri Ouafaa	25
<i>Power Management and Control of a Grid-Connected PV/Battery Hybrid Renewable Energy System</i> Othmani Mohammed, Lamchich My Tahar, Lachguar Nora	32
<i>IT Project Management Models in an Era of Digital Transformation: A Study by Practice</i> Rachida Hassani, Younès El Bouzekri El Idrissi	53
<i>Stability Analysis of a DC Microgrid with Constant Power Load</i> Sarah Ansari, Kamran Iqbal	63
<i>On the Construction of Symmetries and Retaining Lifted Representations in Dynamic Probabilistic Relational Models</i> Nils Finke, Ralf Möller	73
<i>Online Support for Tertiary Mathematics Students in a Blended Learning Environment</i> Mary Ruth Freislich, Alan Bowen-James	94
<i>Solar Energy Assessment, Estimation, and Modelling using Climate Data and Local Environmental Conditions</i> Clement Matasane, Mohamed Tariq Kahn	103
<i>A Study on Novel Hand Hygiene Evaluation System using pix2pix</i> Fumiya Kinoshita, Kosuke Nagano, Gaochao Cui, Miho Yoshii, Hideaki Touyama	112
<i>A Unified Visual Saliency Model for Automatic Image Description Generation for General and Medical Images</i> Sreela Sreekumaran Pillai Remadevi Amma, Sumam Mary Idicula	119
<i>Forensic Analysis of "WhatsApp" Artifacts in Android without Root</i>	127

Mohammad Shadeed, Layth Abu Arram, Majdi Owda	
<i>An Interdisciplinary Approach to Fracture of Solids from the Standpoint of Condensed Matter Physics</i>	133
Mark Petrov	
<i>Interpretable Rules Using Inductive Logic Programming Explaining Machine Learning Models: Case Study of Subclinical Mastitis Detection for Dairy Cows</i>	143
Haruka Motohashi, Hayato Ohwada	
<i>COVIDFREE App: The User-Enabling Contact Prevention Application: A Review</i>	149
Edgard Musafiri Mimo, Troy McDaniel, Jeremie Biringanine Ruvunangiza	
<i>Leakage-abuse Attacks Against Forward Private Searchable Symmetric Encryption</i>	156
Khosro Salmani, Ken Barker	
<i>Digital Competencies of Saudi University Graduates Towards Digital Society: The Case of The University of Tabuk</i>	171
Inam Abousaber	
<i>Cloud-Based Hierarchical Consortium Blockchain Networks for Timely Publication and Efficient Retrieval of Electronic Health Records</i>	179
Alvin Thamrin, Haiping Xu, Rui Ming	
<i>Towards a Framework for Organizational Transformation through Strategic Design Implementation</i>	191
Lynne Whelan, Louise Kiernan, Kellie Morrissey, Niall Deloughry	

Automated Hydroponic System using Wireless Sensor Networks

Mahdi Musa¹, Audu Mabu¹, Falmata Modu¹, Adam Adam^{*1}, Farouq Aliyu²

¹Department of Computer Science, Yobe State University, 620242, Nigeria

²Department of Computer Engineering, King Fahd University of Petroleum and Minerals, 34464, Saudi Arabia

ARTICLE INFO

Article history:

Received: 17 December, 2021

Accepted: 22 February, 2022

Online: 09 March, 2022

Keywords:

Hydroponics

Wireless Sensor Networks

Energy Harvesting

Data Aggregation

OMNET++

ABSTRACT

Researchers have associated agriculture and food processing with adverse environmental impacts like; falls in the underground freshwater table, energy consumption, and high carbon emission. These factors have the worst effect on developing countries. Therefore, there is a need for on-demand food production techniques that require minimum resource utilization. For these reasons, scientists are now focusing their attention on hydroponics. Hydroponics is the process of growing crops without the use of soil. However, different components of the system need to be closely monitored and controlled. In this paper, we compared the performance of an automated hydroponic system using cluster-based wireless sensor networks against a multihop-based one. We used Simponics for the simulation. It is a simulator based on the OMNET++ framework. Simulation results show that both latency and energy overhead of the multihop network increases with the number of nodes. However, they stay constant on a cluster-based network.

1 Background

Hydroponics is the soil-less process of crop cultivation [1]. Modern hydroponic systems were dated back to 1627, albeit knowledge of plants' essential nutrients has not yet fully developed [2]. Nonetheless, scientists have been using this technique to study plants. In 2019, the hydroponic market was approximately worth 8.1 billion US Dollars, and it could reach up to 16 billion US Dollars by 2025 [3]. Hydroponic systems are attractive because of the following advantages [1]: 1) No soil is required. 2) Crops grow faster in this technique than in traditional farming. 3) Neither location nor space is a constraint because nutrients and water are readily available, and plants can be grown vertically [4]. 4) The system also works indoors. Therefore, it is unaffected by adverse climate conditions. 5) Also, pesticides and herbicides are not needed since plants are isolated.

Essentially, a hydroponic system has five main components as shown in Figure 1, these are: 1) the growing area is that area where the plants grow. It uses trays or a network of pipes that deliver nutrients to the plants. But, the plants need adequate spacing for proper growth and protection against infectious diseases. 2) The reservoir is a container that stores the nutrient solution used by the system. Electric pumps transport the nutrient solutions from the reservoir to the growing area. 3) The growing medium is the substrate on which the plants grow [5]. The growing medium substitutes the

soil. It provides support and air to the roots [6]. Some prominent growing mediums are Rockwool, Coconut Fiber/Coconut chips, Grow Rock, and Perlite. 4) Light is necessary for photosynthesis. In indoor hydroponic systems, light-emitting diodes (LEDs) or other light sources of light provide lighting in place of the sun. However, greenhouse-based hydroponic systems use sunlight as a source of light for plants. 5) Finally, the plant that is to be grown is another component of the hydroponic system. In theory, all plants can be grown using hydroponic systems. However, some plants are difficult to grow using hydroponic because of the expenses involved.

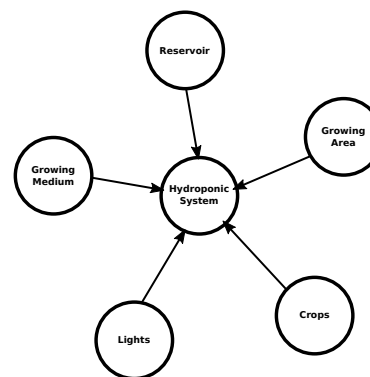


Figure 1: Components of a hydroponic system

*Corresponding Author Adam Adam, Department of Computer Science, Yobe State University, (+234) 806 067 3838 & adam@ysu.edu.sa

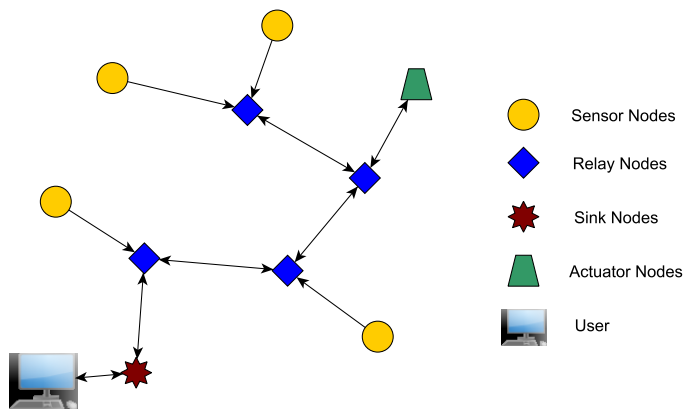


Figure 2: A typical Wireless Sensor Network (WSN)

In large-scale systems, the hydroponic system's components require more maintenance than traditional farming techniques. Automation is necessary to reduce maintenance costs and ensure efficiency. One of the avenues investigated by scientists is the use of wireless sensor networks. A Wireless sensor network (WSN) is a network of limited resources (such as processor, memory, energy) computing devices equipped with sensors for fine-grained sensing of their environment [7]. WSN finds application in precision agriculture, military, health-care, and manufacturing industry [8, 9, 10, 11].

Figure 2 shows a typical WSN. It consists of Sensor nodes, Relay nodes, actuators, Sink nodes, and workstations. The sensor nodes are responsible for sensing the environment. The sensors usually connect to a computing device such as a microcontroller. The microcontroller processes the signals it receives from the sensors. Also, the sensor node has a short-range (typically 100 m) transmitter that sends the sensed data to the sink node. Since the sensor node is short-ranged, there is a possibility that it cannot reach the user. Therefore, a relay node is necessary to forward information to the sink node. The sink node is the gateway that connects the WSN to the outside world. It connects to a workstation or the cloud. The user/network administrator accesses the WSN's data through the workstation for storage and further analysis. The sink node also receives information from the user, such as commands to the actuators in the sensor networks. An actuator converts information into action, thus manipulating the environment. For example, actuators can open or close a valve, turn on or off lights, and the like.

1.1 Motivation

Green Farm [12] is a small desktop hydroponic system designed for growing vegetables at home. The system consists of a growth tray, nutrient reservoir, light-emitting diodes (LEDs) for lighting, and a ventilator fan. The microcontroller controls the lights and fans [13]. Many other similar desktop hydroponic systems are widely available in the market [14, 15, 16]. There are attempts to automate greenhouse-based hydroponics systems: Saaid et al. [17, 18] developed an automatic pH control system for deep water culture (DWC) hydroponics. The pH level of the nutrient solution must be in a certain range to ensure healthy crops. The system uses an Arduino Mega 2560 microcontroller to control the pH level of the system.

The user inputs the appropriate pH level via a keypad. Then, the microcontroller compares the reference pH with the value obtained from the pH sensor. The system uses two tanks, one for acid and another for a base solution. Control valves control the pH level of the nutrient solution in the reservoir.

However, the technique is not scalable because of the large number of wires required to connect the hundreds of sensors for the system. It will also lead to an increase in energy consumption. Moreover, there is redundancy in the data sent to the user, which may overwhelm the user's workstation. These problems can be solved using WSN.

Unfortunately, there are few papers on the deployment of WSN in hydroponics. Figure 3 summarizes the findings of our survey paper [19]. Our findings have shown us that only 13% of the papers we studied are related to WSN. Moreover, none of the papers investigated the impact of scalability and energy harvesting on WSN-based hydroponic systems. We believe it is because the papers use prototypes for their analysis, which makes investigating scalability too expensive and technically difficult. Hence, we develop a simulator that will enable users to investigate energy harvesting and scalability of WSN-based hydroponic systems. Subsection 1.2 discusses the contribution of the paper in details.

1.2 Contributions and Content

In this paper, we proposed a WSN-based automated hydroponic system. It has the following contributions:

1. We developed an automated hydroponic system using clustered WSN. The clustering technique allows the system to scale up by grouping all sensors in one greenhouse under one cluster head (CH).
2. We (to the best of our knowledge) proposed the first use of ultra micro-hydro-turbines (UMHT) [20] for the energy harvesting in WSN applications.
3. Also, a two-level energy harvesting technique is proposed. First, solar energy is harvested from the sun to power the submersible pump, ventilator fan, sensors, and actuators in the reservoir. The sensor nodes carry out the second level of energy harvesting. They scavenge energy from the flow of nutrient solution using UMHT or excess light in the hydroponic system using mini-solar panels [21].
4. We also used data aggregation to reduce energy consumption and transmission delay due to traffic. The sensors measure environmental conditions such as temperature and humidity. These parameters do not vary widely from one location to another. Therefore, a function can be derived to calculate the temperature and humidity of any position in the greenhouse.
5. We also developed an OMNET++ based simulator named "Simponics". It enables the study of hydroponic systems. The modular simulator allows users to investigate the performance of the hydroponic systems when scaled either vertically or horizontally. Furthermore, users can add new sensors, actuators, and transmitters.

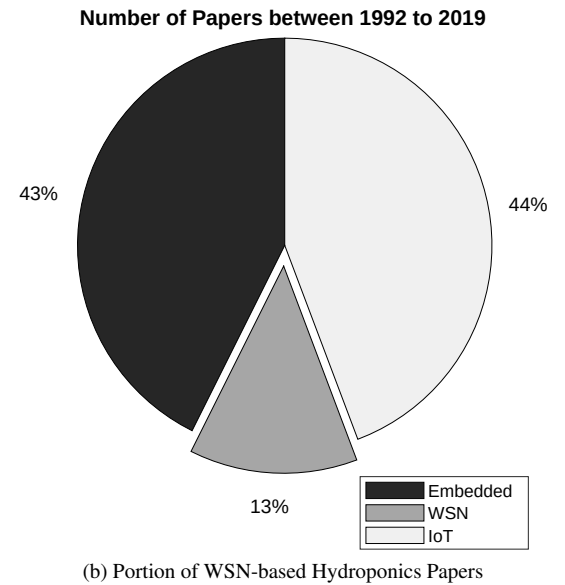
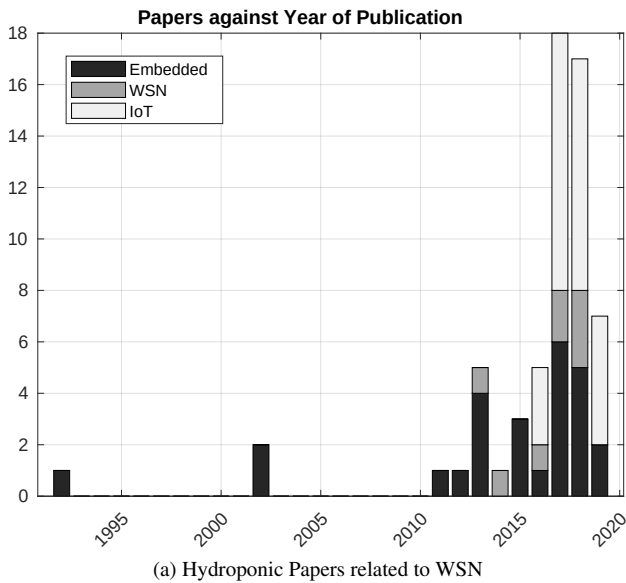


Figure 3: Experimental Setup for the proposed hydroponic system with Two Sensor nodes

The remaining parts of this paper are as follows: Section 2 presents state-of-the-arts hydroponic automation techniques available in both the literature and the industry. Section 3 provides a detailed description of the proposed system. Section 4 describes the experiments and simulations carried out and analyzes the results obtained from the two. Section 5 concludes the paper and presents future works in the research.

2 Literature Review

Figure 1 shows a traditional hydroponic system. It consists of all the five components mentioned in Section 1. The technique for delivering the nutrients' solution to the growth tray determines the type of hydroponic system. Table 1 shows the different types of hydroponic systems currently available. Except for a non-reversed version of a drip system (in which excess nutrient solution in the growth area is allowed to evaporate), the remaining hydroponic systems require nutrients to be renewed or replenished biweekly [22]. However, this is not the only chore needed for a successful hydroponic system. One must ensure that; the pH is between 5.5 and 6.5 [23], the pumps are working, the pipes are not blocked, the nutrient solution is circulating, the optimum temperature is stable, and there is adequate lighting (if the system is indoors) [24]. Furthermore, farmers are not used to these kinds of farming systems. Therefore, they must learn how to farm as well as maintain the hydroponic system.

In a nutshell, hydroponic systems are not as scalable as traditional farming [26] because of the immense care they need. But, researchers noted that automated hydroponic systems minimize system maintenance costs. Thus, increasing its sustainability. Initially, researchers started developing solutions using computer systems: Some authors [27, 28] monitor ions in the nutrient solution. The systems use an array of ion-selective electrodes. However, the nutrient mixture sub-system is not automated. They also ignored the

temperature and PH of the solutions. Both parameters are also necessary for growing healthy plants. Thus, scaling up such systems will be a difficult task.

To reduce energy consumption and improve flexibility, researchers chose embedded systems. As mentioned in Section 1, the researchers in [17, 18], used an Arduino Mega 2560 microcontroller and a pH sensor to monitor the pH level of a Deep Water Culture (DWC) hydroponic system. They then inject acid or base to control the acidity or alkalinity of the nutrient solution using electric valves. In [29, 30] an electric conductivity (EC) sensor, a water temperature sensor, and a water level sensor were added to increase sustainability and efficiency. The level sensor prevents the nutrient solution from spilling out of the reservoir. In [31], the author added a humidity sensor and an LCD screen to improve plant nutrient uptake and system control for the farmer, respectively. In addition, the system allows the farmer to switch between manual to automatic modes. Nalwade and Mote [32] replaced the LCD screen with a GSM module so that farmers get SMS on the status of the hydroponic system.

Alas, the embedded types of automated hydroponic systems are not scalable because of the energy consumption incurred by many sensors and the maze of wires connecting them. Moreover, maintenance and diagnostics difficulty will increase when the number of sensors and microcontrollers increases. To increase the harvest, farmers must scale up the hydroponic system. As such, researchers use the Internet of Things (IoT). IoT is a connection of network-enabled computing devices (i.e., things) to the Internet for data acquisition, communication, and processing [33]. The advantage of using IoT is that it allows the integration of numerous sensors into the system via the Internet.

For example, in [34] the author proposed an Arduino Uno-based monitoring system for an indoor hydroponic system for lettuce, red spinach, and mustard pak choi plants. They used a combination of Open Garden Shield (OGS) and Open Garden Hydroponics (OGH) [35, 36]. The OGS is an Arduino extension board consisting of a

Table 1: Types of Hydroponic Systems in the Literature

S/No.	Hydroponic type	Nutrient delivery	Advantage	Disadvantage
1	Wick	Plant's root take in the nutrients via a wick.	1) The roots have access to both nutrients and air. 2) It is easy to maintained.	The system delivers small quantity of water, which may starve some plants.
2	Deep Water Culture (DWC)	A growth tray floating on nutrient solution on top of which the plants are sown.	1) The system is good for plants that require a lot of water. 2) It is easy to maintain.	1) The plant's root have little or no access to air. 2) It is also prone to fluctuations in pH and nutrient concentration in small-scale systems, which causes health problems for the plants.
3	Ebb and Flow (Flood and Drain)	The nutrient is pumped up to the growth area and excess solution is removed by a pipe just above the accepted level.	Nutrient solution is oxygenated by the solution circulation.	1) Failure of pump may lead to plants dying. 2) System is also known for causing root disease [25].
4	Drip	Plants are fed via a drip system.	1) It can easily be customized to suit the plant grown. 2) There is enough oxygen for the plants' root.	Relatively difficult to manage.
5	Nutrient Film Technique (NFT)	The growth tray is tilted such that the nutrient solution is pumped to one end and drained back to the reservoir at the other.	Nutrient is oxygenated as it circulated.	Relatively difficult to maintain.
6	Aeroponics	Plants' roots are in the reservoir, and they are sprayed with the nutrients.	1) The roots have access to both oxygen and nutrients. 2) Encourages more nutrient uptake in the plants.	Very difficult to manage.
7	Fogponics	Similar to aeroponics but the droplets sprayed are smaller.		

PH sensor, water temperature sensor, Electric Conductivity Sensor, Light Sensor, ambient temperature, GSM / GPRS. A GSM shield sends the data from the sensors to a remote server. For simplicity, small-scale hydroponics come assembled. Moreover, a desktop automated hydroponic system using IoTtalk [37] was developed in [38, 39]. IoTtalk is an IoT platform that manages reconfigurable multi-sensor devices known as MorSensor over the Internet[37]. The systems allow hobbyists to manage hydroponic systems without much experience of how they work. Admittedly, these systems are not suitable for large-scale agriculture.

Therefore, in [40] the author used ThingSpeak. ThingSpeak is like IoTtalk; it is an IoT platform that manages sensors through a remote server in the cloud [41]. The system monitors water level, humidity, and temperature with the help of Raspberry Pi 3 B+ and Field-Programmable Gate Array (FPGA). In [42], the author developed a larger model of the system. The system is modular. Thus, allowing for easy maintenance. However, the system relies on communication with a server that may not be available all the time.

Also, a fixed threshold may not be optimal, since the need and consumption rate of the plants' change as they grow. Since IoT-based hydroponic systems are known to have heterogeneous sub-systems, the researchers in [43, 44] proposed an IoT-based hydroponic system using publish-subscribe middleware over MQTT protocol. The use of middleware allows heterogeneous sub-systems' seamless interoperability. Since the system connects to the Internet, then cybersecurity is essential [45]. Therefore, they integrated TLS/SSL to encrypt MQTT packets.

However, some researchers argue that wireless sensor networks (WSN) are a better solution for developing countries because of their low cost and resilience [48]. The concepts (i.e., WSN and IoT) could be quite confusing, even though this paper has defined both technologies. For clarity, Table 2 compares and contrasts WSN and IoT [46]. The most important part of WSN-based network design is the choice of a suitable transmission protocol that will ensure low latency and energy consumption. One of the most widely used WSN protocols is Zigbee [49]. They developed a Zigbee-based real-time

Table 2: Comparison between WSN and IoT [46].

S/No.	WSN	IoT
1	Network coverage is confined within a local area.	Aided by the Internet, the network coverage is wide.
2	The Network is self-organizing	Nodes either access the network independently or as determined by the backbone.
3	Networks must have a limited number of nodes.	Network supports unlimited number of nodes.
4	Data is only processed by end nodes and aggregator nodes.	Data is processed by all nodes. [‡]
5	Higher energy management capabilities than IoT.	Lower energy management compared to WSN.
6	Low level of heterogeneity.	High level of heterogeneity.

[‡] Recent technologies in IoT allow nodes to connect to a hub for data processing (e.g., Hue Bridge [47])

smart hydroponic system. The system's performance is satisfactory. But, its performance degrades when the sensors are far from the control station.

In [50], the author developed a Zigbee-based aquaponics monitoring system. An aquaponics system is a hydroponic system that gets its nutrient from fish waste in a nearby aquarium [51]. The system consists of Zigbee end nodes and a gateway. Each end node contains a PIC16F877A microcontroller, which collects data from the sensor connected to it. A Zigbee transceiver sends the data to the control station, while the microcontroller uses the data to control the aquaponics through some actuators. For example, the system uses information from the Light Dependent Resistor (LDR) to decide when to switch on the light via a relay switch; it also uses the data from the temperature sensor to deduce when to turn on the cooling fan; lastly, the microcontroller uses the data from the ammonia, chlorine and pH sensors to determine when to turn on the pump or replace the water in the system.

Based on the knowledge we gathered from our survey [19], the WSN monitoring and control systems in the literature are not scalable. They lack energy harvesting, which ensures the sensor nodes do not run out of energy. Also, they lack data aggregation, which reduces redundant traffic in the network. Finally, they lack network clustering, which also reduces traffic.

Granted, IoT systems are scalable. However, they require infrastructure that will enable access to the Internet. Such infrastructure is unavailable in most rural areas in developing countries. Unfortunately, these are where most farmers reside. Therefore, using IoT is not a viable solution in such areas. In rural areas where infrastructure is unavailable, the cost of installation, maintenance (e.g., internet subscription), and system management are arduous if not impossible for the farmers. In 2019 a study was conducted to understand the economic viability of investment in hydroponic systems, especially in emerging countries. The finding suggests that before an investor ventures into hydroponic systems, they need to consider high initial costs. Secondly, he must factor in the expense of monitoring the system a specialist [52]. This paper aims to bridge the gap between these two risk factors through WSN automated hydroponic systems. We do believe that WSN is a viable solution for hydroponics automation in developing countries. Its reliabil-

ity, self-configuration, and recovery from failure make it the right technology for deployment in remote areas [53].

Another study quantifies the effect of desalinated seawater on soil-based cultivation and hydroponic systems [54]. They also investigated their energy consumption and greenhouse gas emissions. They found that the use of NFT increases energy consumption irrespective of the water source. Additionally, there is a significant increase in greenhouse gas emissions compared to soil cultivation. However, the yield is higher [54]. The researchers only compared energy consumption and gas emission of desalinated water in hydroponic systems. Other parameters like temperature and pH were not studied. Thus, our research aimed at bridging this gap. We want to investigate the overhead of sensors in both cluster-based and multihop-based networks.

3 Proposed System

Figure 5 shows the proposed system. We viewed the deployment of the WSN from two perspectives: (1) how the nodes were deployed in the greenhouses as shown in Figure 5a, and (2) how the deployed nodes from the different greenhouses communicate with one another to send data to the workstation (see Figure 5b).

3.1 Greenhouse

Firstly, in each of the greenhouses, the sensor nodes were deployed as shown in Figure 5a. Our proposed system is a modified NFT hydroponic system. We chose NFT because it requires no growing medium, timer, and air pump [19]. Thus, the system consumes less energy while maintaining good plant health. The NFT uses perforated pipes for the growth area. It also uses a light-emitting diode (LED) and a reflector as a light source. Then, a submersible pump delivers the nutrient solution from the reservoir to different parts of the system. A rechargeable battery is used to power the pump and the LED. The system uses solar panels to charge the battery. As shown in Figure 5a, There are two solar panels; primary and secondary. The earlier provide energy to the LEDs and pumps, while the latter connects to the sensor nodes for energy harvesting.

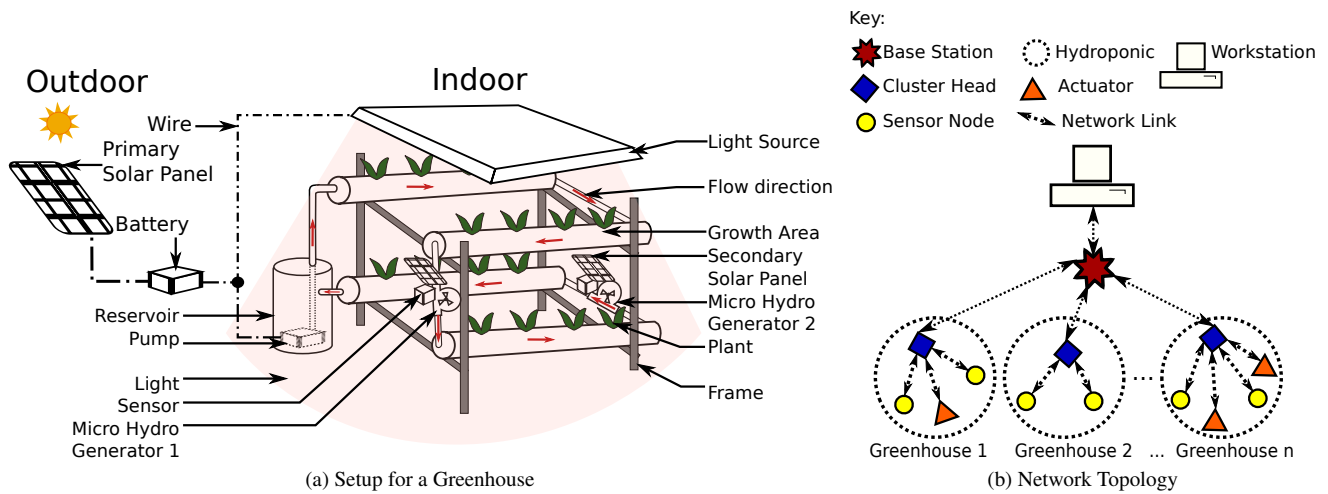


Figure 5: Setup for the proposed hydroponic system

The sensor nodes measure temperature, humidity, electric conductivity (EC), and pH. The sensor nodes use mini-solar panels [21] to scavenge the excess light from the growth area. Only the light that falls on the leaves helps in photosynthesis. Therefore, the light that falls elsewhere is a waste. Hence, the mini-solar panel converts the wasted light energy into electrical energy. Thus, improving the efficiency of the system. We call them secondary solar panels because they feed on the unused light energy drawn from the primary solar panel. Another source of energy wastage is the kinetic energy and the potential energy of the nutrient solution pumped by the submersible pump. We used ultra-micro-hydro turbines (UMHT) [20] to harvest the kinetic and potential energy of the moving fluid. A UMHT is a credit card-sized turbine that converts fluid's mechanical to electrical energy. The sensor nodes sense the environment and forward the data to a base station (BS). The BS then sends the data to a workstation. The workstation serves as the control center where the network administrator can monitor the hydroponic systems.

3.2 Network

Secondly, Figure 5b shows the network topology of the proposed system. In the figure, each of the greenhouses has several WSN nodes deployed. The number and type of nodes depend on the requirement of the greenhouse. A small greenhouse will have fewer sensor nodes. Also, greenhouses that use DWC will have few or no actuators since they do not have pumps and valves. Each greenhouse has a cluster head (CH). The CH is responsible for collecting data from other nodes in that greenhouse and forwarding them to the BS. The BS then sends the data to the workstation.

The sequence diagram in Figure 6 shows how the proposed system's network works. It shows that the sensor nodes go to sleep for a predetermined period (n sec). The sleep period depends on the tolerance of the plants towards changes in the environmental parameters. In [55], the author took measurements every 5 mins. For fine-grain sensing, our proposed system wakes up every 60 s to sense the environment. The system takes 6 s to complete the sensing session because the slowest sensor is the humidity sensor, which takes 6 s to complete its reading [56]. Whenever they wake up, the

nodes will sense the environment and forward the results to the CH. The CH gathers the data from the different sensor nodes, aggregates them, and forwards the aggregated data to the BS. Note that the CH and the BS do not go to sleep to avoid data loss.

$$s = \sum_{i=1}^n x_i \quad (1)$$

$$ss = \sum_{i=1}^n x_i^2 \quad (2)$$

$$\mu = \frac{s}{n} \quad (3)$$

$$\sigma = \sqrt{\frac{ss}{(n-1)} - \mu^2} \quad (4)$$

The CH carries out the Data aggregation. Data aggregation is a process for gathering data and summarizing it to reduce data redundancy [57]. For data aggregation of the temperature and humidity, the system uses mean and standard deviation as shown in Equation 3 and 4, respectively. We chose the mean and the standard deviation because we are interested in how the temperature and humidity vary across the greenhouse. The pH and EC are only measured at the nutrient reservoir because the movement of the solution through the system keeps the nutrient composition homogeneous. Therefore, there is no need for data aggregation for these data. For a given environmental parameter like temperature, the system uses Equation 1 and 2 to calculate the sum and the sum of squares, respectively. After the summation of all temperature values from all the sensor nodes, the mean μ and standard deviation σ are calculated according

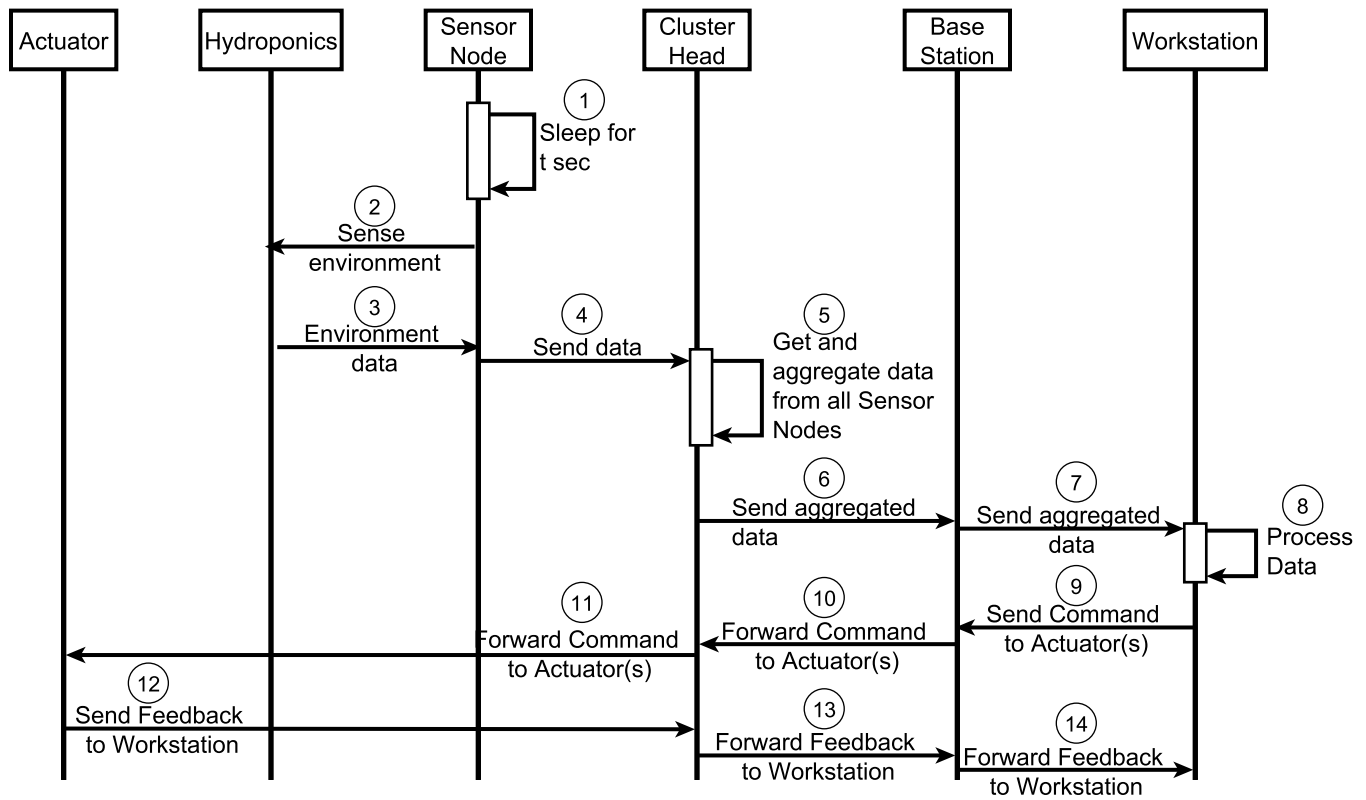


Figure 6: Sequence Diagram of the Proposed System

to Equation 3 and 4 respectively.

Algorithm 1: Data aggregation at the CH

Data: Readings from all Sensor nodes

Result: Aggregated data

```

1  $y \leftarrow 0$  ;
2  $n \leftarrow$  Number of children ;
3 while ( $y < n$ ) do
4    $x_{i,j} \leftarrow$  data  $i$  from sensor node  $j$  ;
5    $y \leftarrow y + 1$  ;
6   for (Environmental Parameters  $i$  in  $x_j$ ) do
7      $s_i \leftarrow s_i + x_{i,j}$  ;
8      $ss_i \leftarrow ss_i + x_{i,j}^2$  ;
9   end
10 end
11 for (Environmental Parameters  $i$  in  $x_j$ ) do
12    $\mu_i = s_i/n$  ;
13    $\sigma_i = \sqrt{(ss_i/(n - 1)) - \mu_i^2}$  ;
14 end

```

Algorithm 1 describes how the CH aggregates data it obtains from its children. Each of the j sensor nodes in a greenhouse measures i different environmental parameters. The sensor nodes wake up every 60 s and sense the environment. Then the data they obtain is forwarded to the CH. Line 6 shows that the CH extracts the environmental parameters from each packet it receives. Each of the parameters' values and the square of the values are separately accumulated with the respective parameters of the other sensor nodes as

shown inline 7 and 8 respectively. For example, if the temperature is the first parameter, then s_1 and ss_1 are the accumulated sum and sum of squares of the temperature from all j sensor nodes. Then the accumulated sum and sum of squares is s_2 and ss_2 if humidity is the second parameter. The same goes for the remaining i parameters. The CH knows the number of its children (n) in the cluster. Thus, it knows when it has gathered all the data from all sensors (see inline 3). Then, the CH uses the value of n to calculate μ and σ of each parameter from s and ss . However, this is only possible when the CH receives all the data from the nodes in its cluster. As shown in line 12 and 13, μ and σ are calculated for each of the i parameters using Equation 3 and 4, respectively. In the end, the CH obtains i number of means and standard deviations. The CH then forwards them to the BS, who then forwards them to the workstation.

The CH sends the aggregated data to the BS. The BS then forwards the aggregated data to the workstation. The received data is processed, and the system uses the results to control the environment. For example, if the temperature is high, the system sends a command to the actuator(s) to start the cooling fans to improve air circulation and lower the temperature. As shown in Figure 6, the workstation sends instructions to the controllers through the BS and the CH. Whenever an actuator receives a command, it responds with its state (e.g., on/off or open/close) as feedback to the workstation. It allows the user/administrator to know what is happening in the given greenhouse. The actuator sends its response to the workstation through the CH and then the BS. Currently, our proposed simulator has no actuators. However, they can be added by users, depending

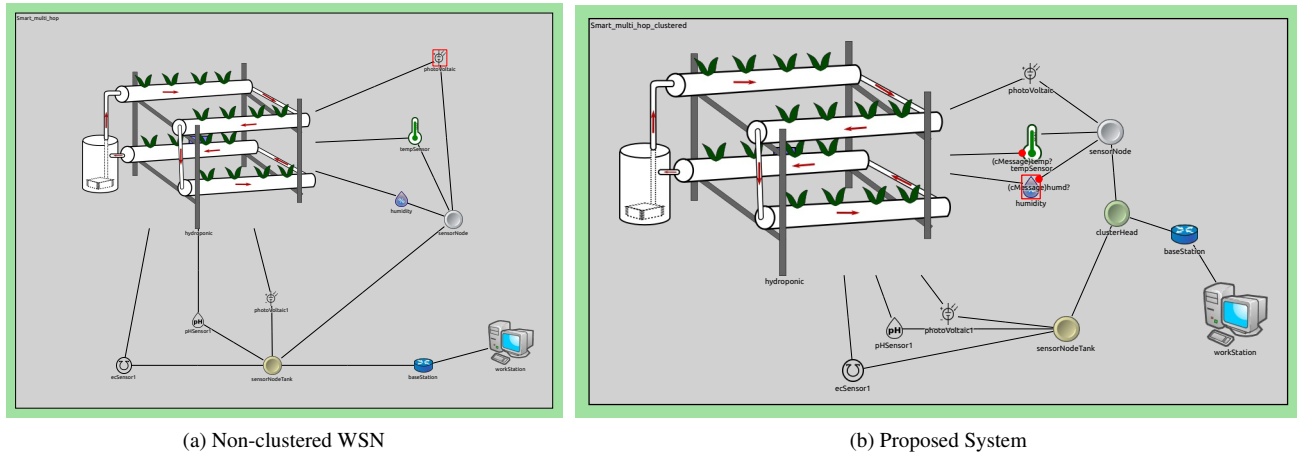


Figure 7: Experimental Setup for the proposed hydroponic system with Two Sensor nodes

on the scenario.

4 Discussion of Results

Simponics++ is an OMNET++ based simulator we developed to investigate the performance of our proposed hydroponic system. OMNeT++ is an object-oriented library and framework for developing a discrete event- and a modular network simulation based on C++ programming language [58]. The word “network” means any system that consists of the interconnection of sub-systems. Examples of such networks are; queuing networks, communication networks, and on-chip networks. The OMNET++ framework is modular, which means that simulators can be developed by modeling the sub-systems independently and then connecting them like LEGO blocks to form a complete system. Therefore, users can easily modify our simulator to add, edit, or delete components (or sub-systems) such as growth areas, sensors, actuators, transceivers, and others. Simponics++ is available on GitHub for modification and this experiment’s reproduction [59].

Figure 7 shows the experimental setup for the proposed system. On the left, in Figure 7a, is a typical non-clustered WSN. It consists of two sensor nodes; the one at the top reads the temperature and humidity of the greenhouse, while the other (at the bottom) is connected to the nutrient reservoir to measure the pH and electric conductivity (EC) of the nutrient solution. The network uses the shortest path first technique to transmit data to the base station (BS): the sensor at the top of the diagram sends its data to the node at the bottom because it is closer to the BS. This node then sends the data to the BS on behalf of the sensor node at the top. We use the non-clustered network as a control experiment to compare with the proposed system. The performances investigated are the energy consumption and latency of the network. The energy consumption rate of the components determines how long it monitors the hydroponic system without the network shutting down. The latency of the network determines the agility of the system in reporting and responding to changes in the environment, such as pH, temperature, light, and nutrient availability.

On the right is Figure 7b, which shows the proposed system.

The figure shows a cluster-based WSN with two sensor nodes: the one at the top reads the temperature and humidity of the greenhouse, while the other (at the bottom) is connected to the nutrient reservoir to measure the pH and electric conductivity (EC) of the nutrient solution. Both sensor nodes take readings of the environment. They send the data directly to the cluster head (CH). The CH then aggregates the data from both of the children before sending the data to the BS. The BS then sends the reading to the workstation for analysis, as explained in Section 3.2.

$$R_{th} = 71,829.88e^{(-0.0365T)} \quad -55 \leq T \leq 150 \quad (5)$$

Table 3 shows the simulation parameters used in all the simulations in this paper. The table’s columns show; the name of the variables as coded in the simulator, the value for the variables, a description of the variables, and the device or sensor to whom they belong. Also, we have provided all the simulations carried out in this paper on GitHub [59]. We set parameters for the different sensors from their respective datasheets. The transceiver is a Zigbee sensor mote whose characteristics are in [60]. The thermistor is a variable resistor whose resistance changes with the temperature. In this paper, the thermistor is NTC 10k [66]. We derive Equation 5 from the table in [66]. It determines the resistance of the thermistor (R_{th}) when given its temperature (T).

$$f_s(t) = \begin{cases} -25.6t^2 + 614.4t - 2764.8 & 6 \leq t \leq 18 \\ 0 & \text{elsewhere} \end{cases} \quad (6)$$

Concerning the photovoltaic cells, there is a need for solar irradiance from which the cells will harvest energy. We model the solar irradiance in this simulation from the month of October in Nigeria, as reported in [67]. From the paper [67], Equation 6 is developed, where t is time in hours in the 24-hour format and $f_s(t)$ is the solar irradiance in W/m^2 .

We carried out four experiments to investigate the energy and latency performances of the proposed system. Figure 8 shows the topologies of the experiments. The experiments were set up in Simponics++ as follows:

Table 3: Parameters for the Simulation

S/No.	Name	Value	Description	Device/Sensor
1	I_{tx}	32.0 mA	Transmission current	Transceiver [60]
2	I_{rx}	25.5 mA	Reception current	
3	V_{trx}	3.3 v	Transceivers voltage	
4	rate	250 kbps	Transmission rate	
5	pktSize	1500 B	Packet size in bytes	
6	battery	3	number of AA size batteries	
7	mu	2	inter-arrival rate	
8	voltage	1.7 v	Voltage of battery AA	
9	current	2750 mAh	Battery Current	
10	V	5 v	Supply voltage for sensor	Electric Conductivity [61, 62]
11	I	50.0 mA	Sensing current (A)	
12	T	1 s	Minimum reading time is 1s.	
13	I_{sleep}	0.7 mA	Current consumption during sleep	
14	V	5	voltage	Humidity Sensor [56]
15	I	2.5 mA	Sensing current (A) consumed	
16	T	6 s	Minimum reading time is 6s.	
17	I_{sleep}	100 μ A	Current consumption during sleep	
18	V	6	Supply voltage	Photo voltaic Cell [63]
19	I	170 mA	current	
20	T	1 s	time seconds	
21	L	125 mm	Length	
22	W	63 mm	Width	pH Sensor [64, 65]
23	V	5	Supply voltage	
24	I	14 mA	current	
25	T	1 s	time seconds	
26	I_{sleep}	2 mA	Current consumption during in-activity	Thermistor
27	v	5.0 v	Supply voltage	
28	R	10 k Ω	Resistor in the potential divider	
29	v	5.0 v	Generated voltage	Ultra Micro Hydro Turbine [20]
30	I	150 mA Ω	Generated current (A)	
31	τ	60 s	Sampling Time	Sensor Nodes

1. A two (2) node non-clustered WSN developed. One of the sensor nodes (SN) senses temperature and humidity and sends the data to the BS through the other sensor node (SNT). The BS then sends the data to the workstation. The sensor node connected to the tank is known as SNT. It senses the pH and electric conductivity (EC). It forwards sensed data to the BS, which then sends it to the workstation. Figure 8a shows the topology of the experimental setup.
2. In the second experiment, two more sensor nodes; SN1 and SN2. Both sensor nodes added are of SN type, meaning that they sense temperature and humidity. Figure 8b shows the topology of the setup for this experiment: SN2 sends its data to the BS through SN1, then SNT, while the SN1 sends its data to the BS via SNT. However, SN forwards its data to the BS in one hop. Whenever the BS receives a packet, it forwards it to the workstation immediately for data processing.
3. The third experiment is the proposed system. The system consists of two sensor nodes SN and SNT, similar to the ones

in (1) above. Figure 8c shows the setup. Both sensor nodes send their data to the cluster head (CH). The CH waits until it gathers and aggregates all information. Then it forwards it to the BS, which then forwards it to the workstation for further processing.

4. In the final experiment, the number of nodes in the clustered network of the proposed system was increased to four, as shown in Figure 8d. The network consists of three sensor nodes of type SN, labeled SN, SN1, and SN2. Also, there is one SNT sensor node connected to the tank to sense pH and EC, like in the experiment (3) above. All nodes send their data to the CH. The CH aggregates all the data into one packet and sends it to the BS. Finally, the BS sends it to the workstation for data analysis.

We repeated the experiments with solar panels installed in all nodes (except the workstation). The solar panels installed on the nodes are secondary (small size). They scavenge the light energy that missed the leaves.

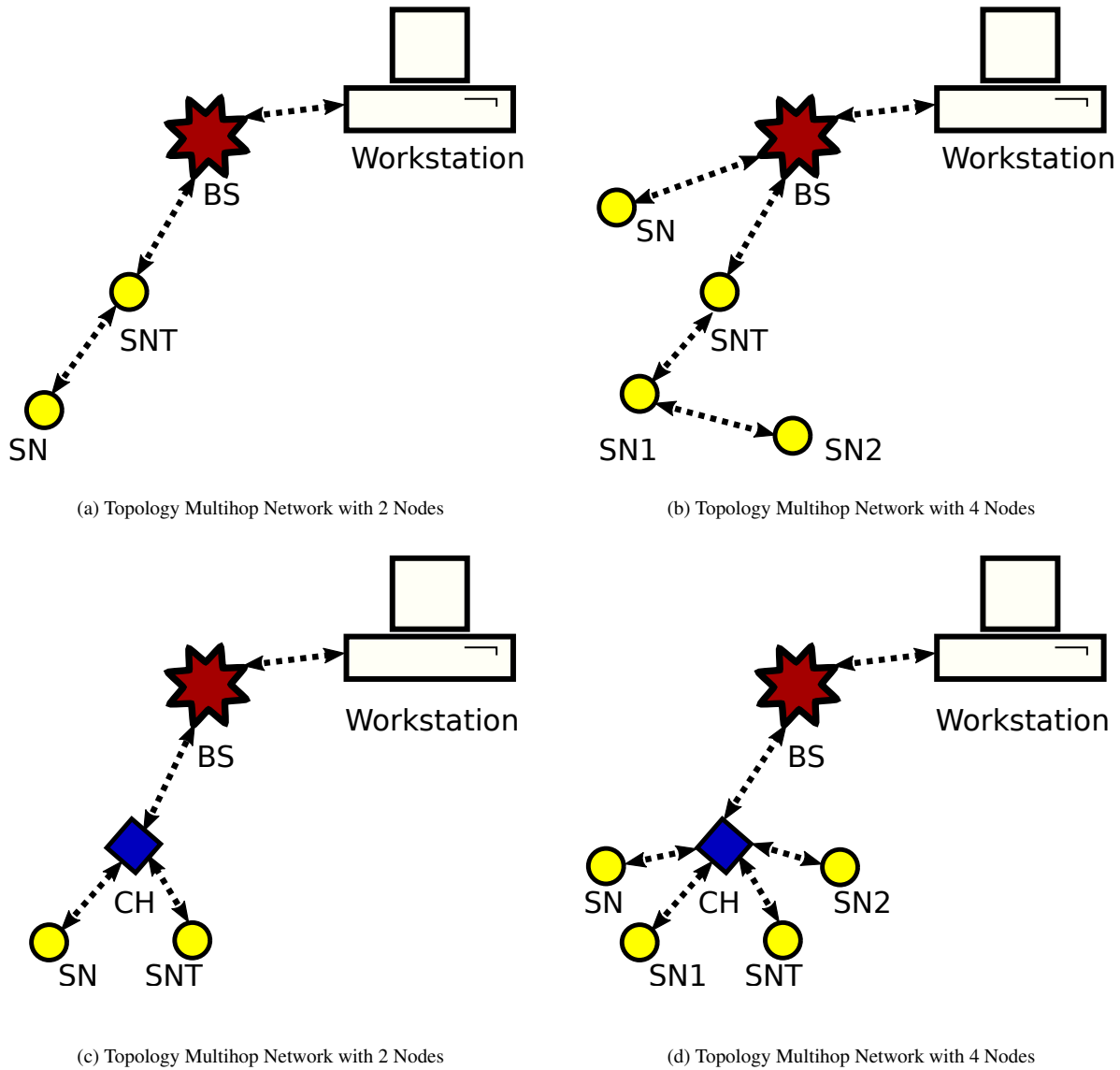


Figure 8: Energy of a Multihop over a Period of 30 Days

4.1 Energy Consumption

Studying the energy consumption of a WSN is very important. It allows us to know how long the system lasts. There is no network without energy. We compare the energy consumption of the proposed system that of a multihop non-clustered WSN with the same number of sensor nodes. The multihop WSN uses the shortest path from any node to the BS. In this paper, we investigate the performance of the proposed system (in comparison with the multihop non-cluster WSN) as the number of nodes increase.

$$E(t) \approx E_{total} - t \left(\frac{E_s - E_{ss} + (1 + 2N)E_{trx}}{\tau} + P_{ss} \right) \quad (7)$$

$$\text{where } E_s = \sum_{i=1}^{sn} E_i, \quad (8)$$

$$E_{tx} \approx E_{rx} = E_{trx},$$

$$\text{If } \alpha = \frac{E_s}{E_{trx}} \quad (9)$$

$$\text{and } \beta = \frac{E_{ss}}{E_{trx}} \quad (10)$$

$$\text{Hence } E(t) \approx E_{total} - t \left(\frac{(\alpha - \beta + (1 + 2N))}{\tau} E_{trx} + P_{ss} \right) \quad (11)$$

Figure 9 shows the results for energy consumption obtained from the experiments without energy harvesting. The following observations are vivid: First, the relationship between the energy consumption of any node with time in any configuration is linear. Equation 7 corroborates it. In the equation, $E(t)$ is the instantaneous energy consumption at the time (t). E_{total} is the initial energy of the battery when fully charged. E_s is the sum of energy consumption of each of the sn sensors connected to the node as described by Equation 8. E_{tx} is the energy during transmission of data, and E_{rx} is the energy consumption due to packets received from the N sensor nodes. Equation 11 is a simplification of Equation 7. It

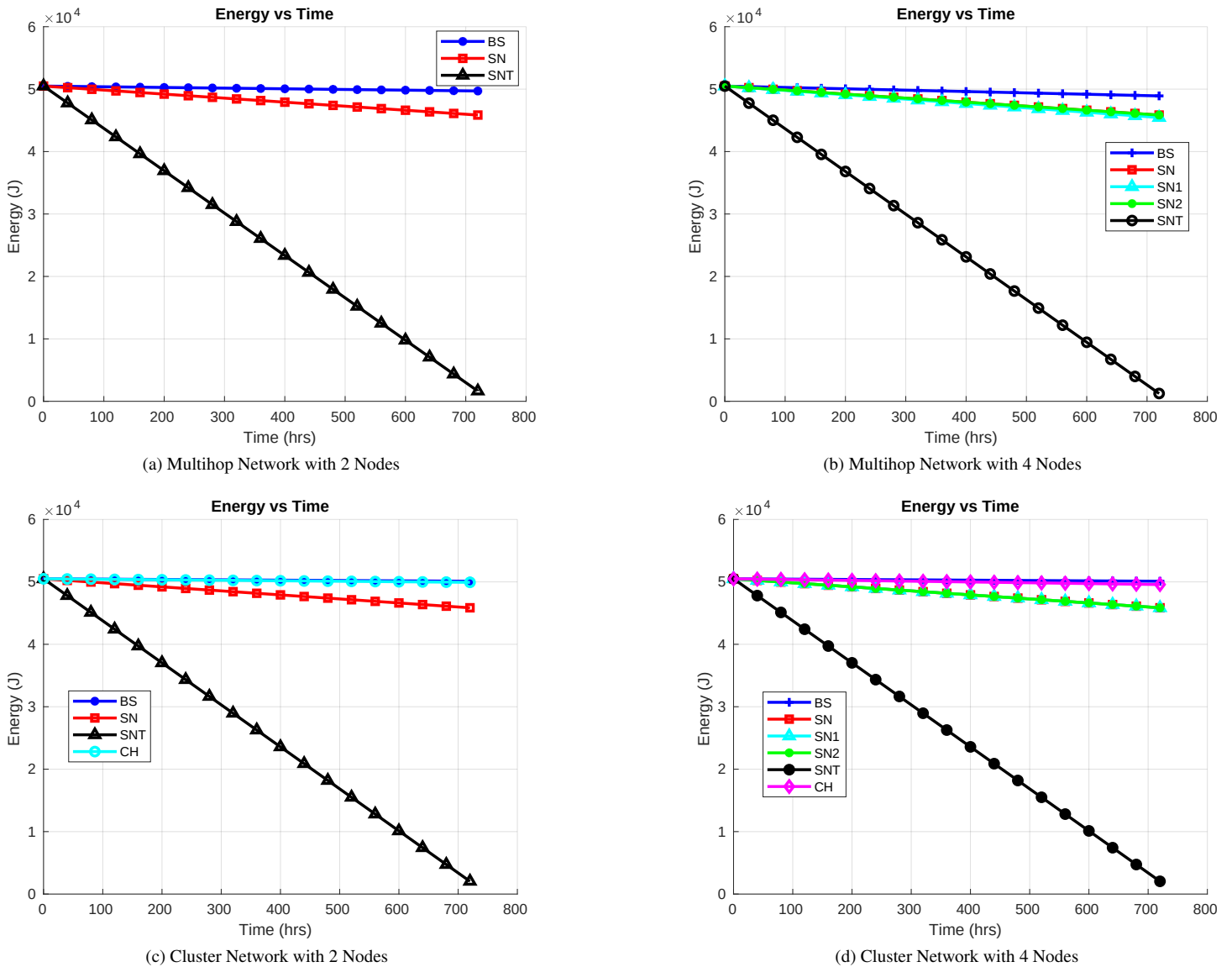
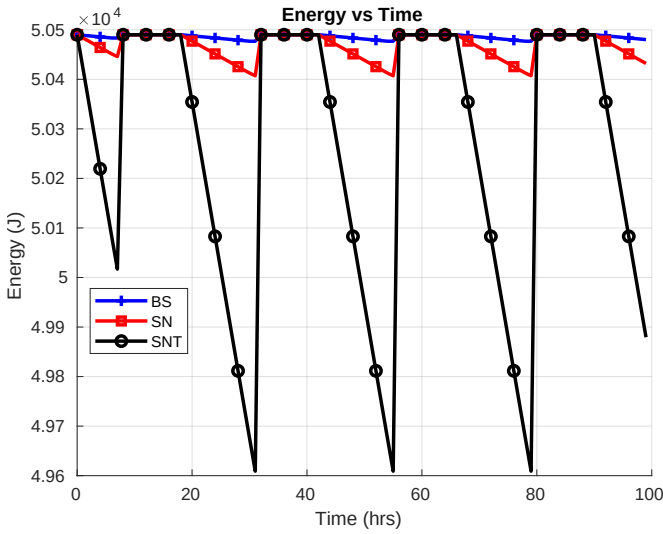


Figure 9: Energy of the Proposed System over a Period of 30 Days

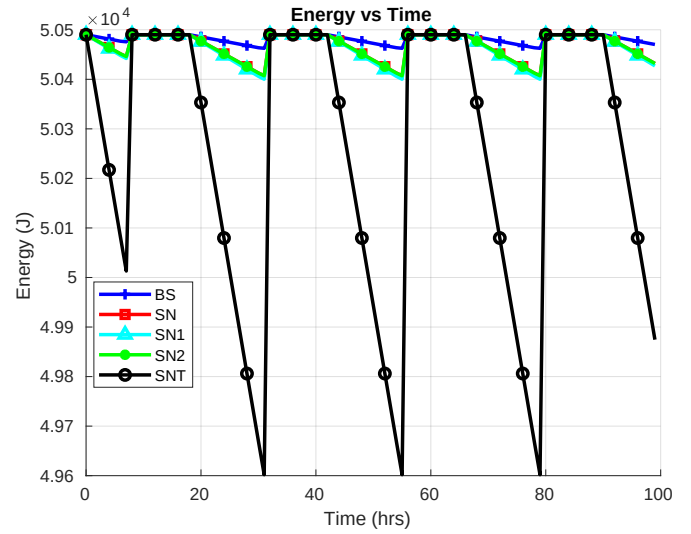
supports the simulation results that argue the relationship between energy and time is approximately linear. The derivation for this model is available in Appendix 6. Also, the model agrees with the simulation on the y-intercept being E_{total} and slope determined by $(2N + \alpha + 1)E_{trx}/\tau + P_{ss}$. The slope is the energy consumption rate and is proportional to the number of children nodes, the energy consumption of the sensors in the node (be it sleeping or sensing). Also, it is inversely proportional to the sampling period τ . In general, the energy consumption for the sensor nodes in the clustered network is less because $N = 0$, meaning no sensor node has the overhead of forwarding another sensor node's data. Therefore, we are left with $(\alpha - \beta + 1)E_{trx}/\tau + P_{ss}$. The reduction is not well pronounced because of $\alpha \gg N$. It means that the sensor nodes consume the same energy regardless of the size of the network in the clustered network, as opposed to the non-clustered network, where the energy consumption rate of any given sensor nodes increases by $2NE_{trx}/\tau$ whenever the number of children (N) for that node increases.

Second, the energy consumption of the sensor node at the tank (i.e., SNT) is by far more than the energy consumption of any other node in any case. It is due to the energy consumption of the EC sensor, which consumes 250.0 mJ of energy when sensing data from the environment, compared to the 4.25 mJ of energy consumed by the humidity and temperature sensors (combined). To accurately verify this claim, we compare the energy consumption rate of SN and SNT in the case of the clustered network, since they are $N = 0$. From Figure 9c, we obtain the slope of SNT and SN. We get $0.0185 J_s^{-1}$ for the SNT and $0.0013 J_s^{-1}$ for SN. It means that the SNT sensor consumes 14 times more energy than SN. Therefore, low power sensor nodes are necessary for a sustainable and scalable smart hydroponic system.

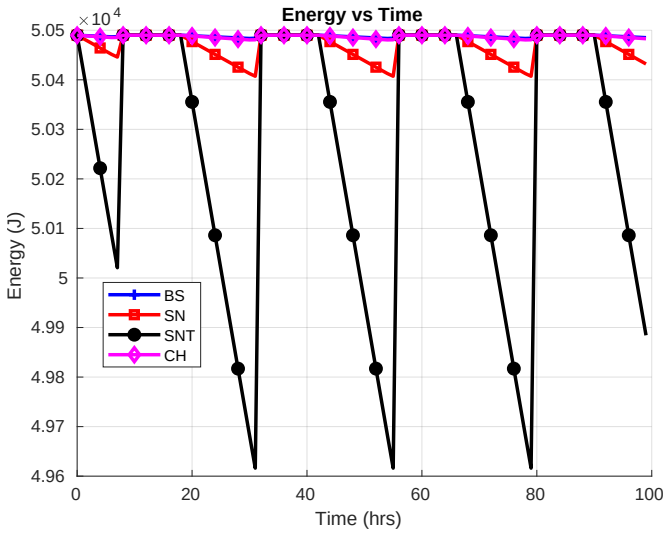
The third observation is that if the same system changes from non-clustered to clustered, the energy consumption at the BS drastically falls. Thus, extending the life of the network. Note that if the BS is dead, the network becomes useless. The clustered net-



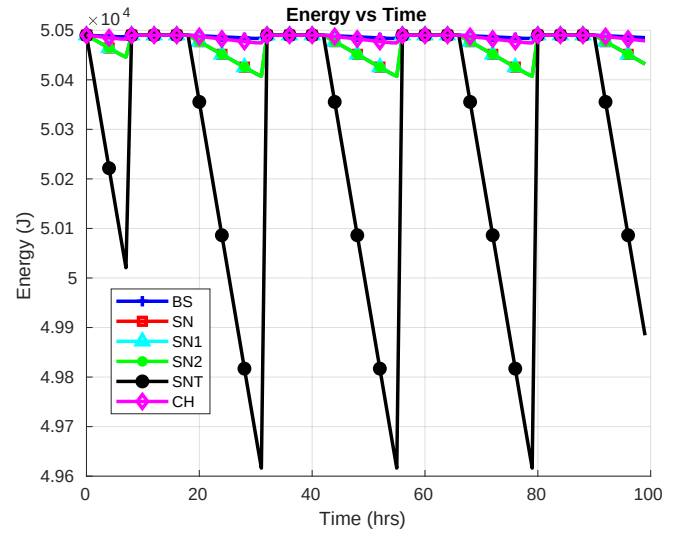
(a) Multihop Network with 2 Nodes



(b) Multihop Network with 4 Nodes



(c) Cluster Network with 2 Nodes



(d) Cluster Network with 4 Nodes

Figure 10: Energy Consumption Sensor Nodes with Energy Harvesting over a Period of 30 Days

work saves energy because data aggregation reduces the number of communications between the BS and the other nodes.

The energy consumption experiment is repeated with energy harvesting as shown in Figure 10; we added a 125 mm by 63 mm solar panel to the sensor nodes, cluster heads (CHs), and the base stations (BSs). The solar irradiance is modeled after the solar energy available in Nigeria in the month of October [67]. The experiment tries to answer the question, “Is energy harvesting going to sustain the networks?”. For the solar panel to suffice, it must satisfy Equation 12, where $f_s(t)$ is the solar energy per unit area produced during the day, A is the surface area, and $E(t)$ is the energy consumed in 24 hours (i.e., day and night) by the sensor node. Since the area of the solar panel is $7.875 \times 10^{-3} \text{ m}^2$. Then the sensor node must not consume more than 58.06 Wh or approximately 209.02 kJ per day as expressed in Equation 12.

$$\int_6^{18} f_s(t) dt \times A > E(t) \quad \text{at } t = 24 \quad (12)$$

$$\implies E(24) < 58.06 \text{ Wh} \quad (13)$$

Clearly, Figure 10 shows that the solar panels quickly replenish the energy consumption of both clustered and non-clustered networks. It is because the energy consumed by the sensor nodes does not exceed 209.02 kJ daily. The daily energy consumption of the SNT, which has the most energy consumption, is only 900 J daily. Thus, the solar panel replenishes energy consumed as early as the first two hours after sunrise. Then, the sensor nodes consume energy directly from the solar panels, never to use the battery again until sunset. Then the solar panel replenishes it in the morning. This consume-replenish cycle is evident in the figures. However, in all sub-figures in Figure 10, the energy consumption of the first cycle is followed by subsequent identical cycles. It is because the simulation

starts at 0000 hours in the morning. Thus, in the first cycle, the sensor nodes use the battery energy for six hours only. Then, it is quickly replenished two hours after sunrise. The following day the sun sets at 1800 hours. Hence, the nodes use the battery for 12 hours rather than the six hours the previous day. Thus, the sudden increase in energy consumption of the battery in other days after the first.

$$\begin{aligned} \Delta E(t) &\approx E_{\text{multihop}}(t) - E_{\text{clustered}}(t) \\ &\approx \frac{2(N - CH)}{\tau} E_{\text{trx}} \\ &\approx \frac{2N(1 - CH/N)}{\tau} E_{\text{trx}} \end{aligned} \quad (14)$$

Figure 11 zooms in on the waveform for 24 hours. We chose the second day because the energy consumption is more than the first, as we have seen earlier. The sub-figures are in two columns. The left row shows the behavior of sensor nodes in the multihop network (both 2 and 4 sensor node networks), and the right column shows the behavior of the clustered network. In the first row, Figure 11a and 11b show the energy consumption of the BSs in the different experiments for the multihop and the clustered network, respectively. Figure 11a shows that the energy consumption increases with an increase in the number of sensor nodes. But, in the clustered network in Figure 11b, the energy consumption is independent of the number of sensor nodes in the network. This observation is in line with the prediction of Equation 11. All the BSs in the clustered networks consume equal and less energy because the CHs behave like their regional BS. By gathering data and aggregating it, the CH reduces N several transmissions to one. Thus, reducing the workload of the BSs. Equation 14 shows an approximation of the saved energy. It shows that one can save more energy by putting more nodes in a cluster while keeping a low number of CHs. However, the cluster heads consume almost the same amount of energy as the BS of the non-clustered (multihop) network because the CH manages the same number of nodes as the BS in the non-clustered network. Nevertheless, the clustered network continues to work even if a CH is dead. In some implementations, the sensor nodes vote for another CH.

The energy consumption of the sensor nodes is the same in both cases as shown in Figure 11c and 11d, except for the sensor node SN1. SN1 consumes more energy than the others because of the overhead due to its child SN2. This instance shows that the clustered network helps the sensor nodes in the network to conserve energy since they do not have any children to incur overhead on them. However, it is noteworthy that the energy overhead due to forwarding data (receiving and transmitting to the next hop) is approximately 9.11 mJ. It is less than the energy consumption of the sensors. It is evident in Figure 11e and 11f; even though the SNT nodes in the multihop network have children, their overhead is not evident because of the energy overhead of the EC sensor. Thus, the two figures look identical despite the packets the SNT sensor nodes forward. In a nutshell, all the sensor nodes replenish their energy and switch to using the solar panel for at least 10 hours before switching back to the battery daily.

4.2 Latency

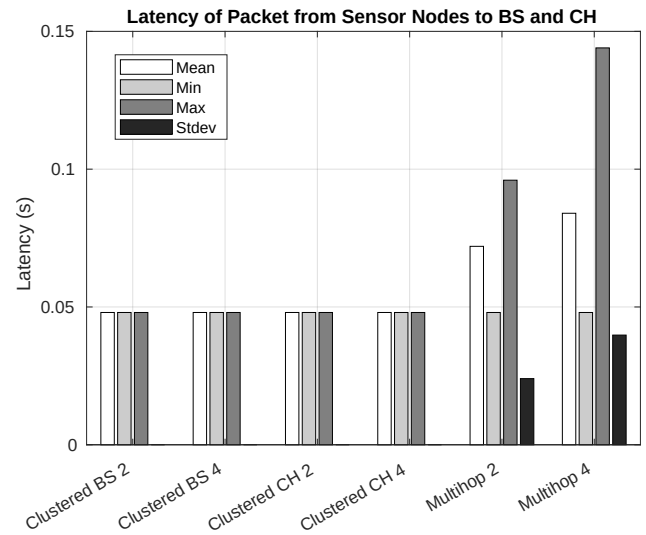


Figure 12: Latency

$$\lambda_{\text{cluster}} = \lambda_{SN_CH} + \lambda_{CH_BS} \quad (15)$$

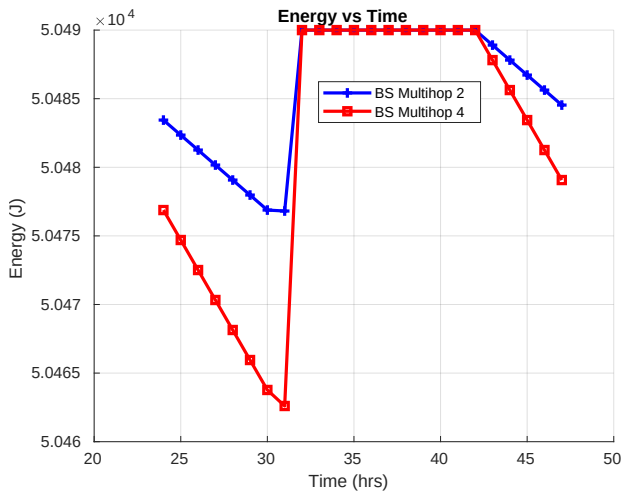
$$\lambda_{\text{multihop}} = x \times \lambda \quad (16)$$

where $\{x : x \text{ is longest routes in the network}\}$

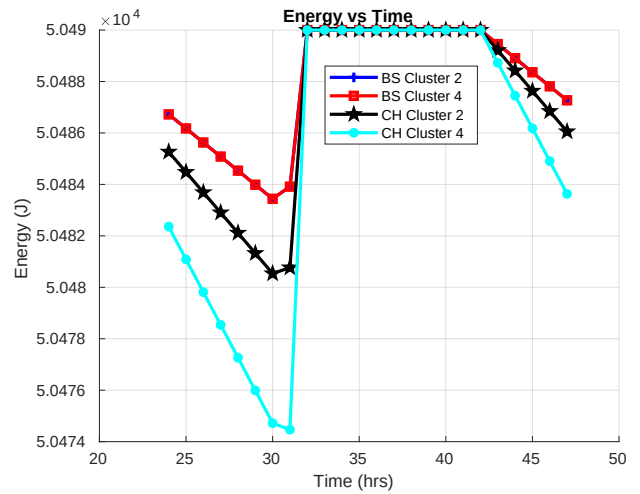
Another matter of great importance in WSN-based hydroponic systems is the latency of the system — how fast does the network report an event? Latency is the time it takes data to reach the BS from a sensor node. Monitoring plants' health is a time-sensitive matter. Therefore, the system's latency provides an insight into the scalability of both the cluster-based and the multihop-based WSN. We carry out some experiments to investigate the latency of the two systems. Figure 12 shows the latency of the different experimental setup: We label the latency of the CH to BS in a two sensor node network as "Clustered BS 2". Therefore, "Clustered BS 4" means CH to BS latency of a clustered network with four sensor nodes in the cluster. Similarly, the latency of sensor nodes to CH in two sensor nodes per cluster and four sensor nodes per cluster are labeled "Clustered CH 2" and "Clustered CH 4", respectively.

Regarding the latency of the multihop network, we label them as; "Multihop 2" and "Multihop 4", which represent the latency of a multihop network with 2 and 4 hop networks, respectively. In the multihop network, the maximum latency depends on the latency of the longest route to the BS as shown in Equation 16. However, the latency of the clustered network remains the same. The sensor nodes in the network forward its data to its CH, who then forwards the packet to the BS. Therefore, the latency of a clustered network is equivalent to the time it takes data to move two hops, as shown in Equation 15.

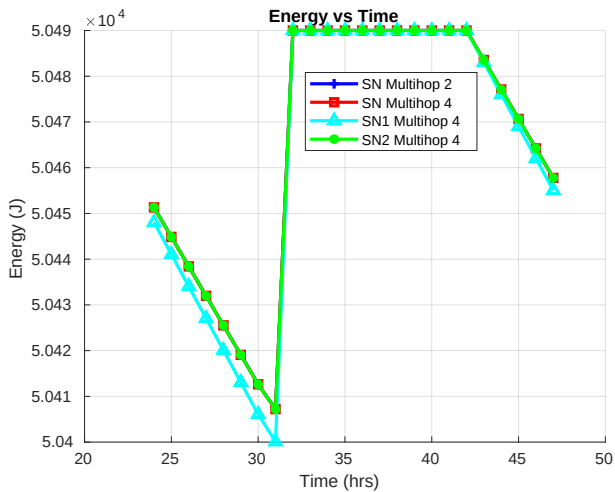
Thus, the clustered network is more scalable since the latency remains the same with an increase in the number of sensor nodes. Nonetheless, the latency of the multihop network increases linearly with an increase in the size of the longest route. Therefore, the latency can only become high in the case of an industrial-scale automated hydroponic system with thousands of sensor nodes. Similarly,



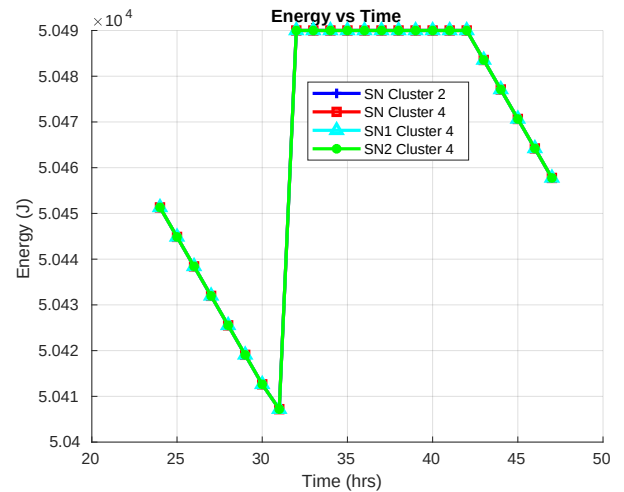
(a) Multihop BS Energy Consumption



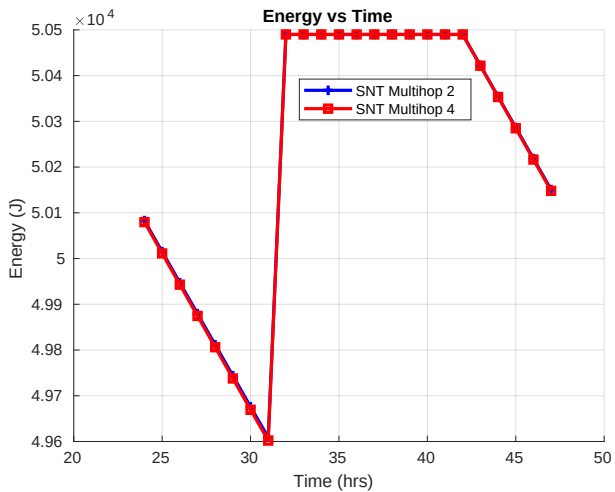
(b) Clustered BS & CH Energy Consumption



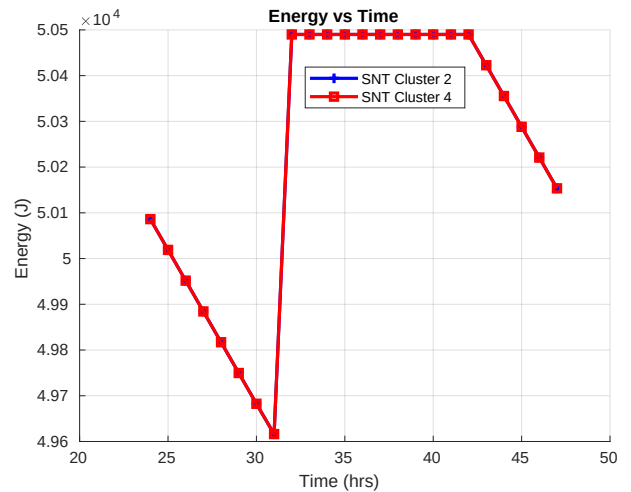
(c) Multihop SN Energy Consumption



(d) Clustered SN Energy Consumption



(e) Multihop SNT Energy Consumption



(f) Clustered SNT Energy Consumption

Figure 11: Energy Consumption Sensor Nodes with Energy Harvesting over a Period of 30 Days

the energy consumption of the sensor nodes in the clustered network is constant regardless of the number of nodes in the network. But, the CH's energy consumption increases linearly with the number of nodes. Since there are few CH in a network, one can solve their

energy overhead problem by providing more energy sources. However, for the multihop network, the energy consumption increases for all nodes that have children. Therefore, it is too expensive to provide more energy sources to the sensor nodes while maintaining portability.

5 Conclusion and Future Works

In this paper, we simulated both cluster-based and multihop WSN on an NFT hydroponic system. Within the limits of experimental errors, we were able to conclude that, clustered network helps to develop a scalable hydroponic system because the clustered network saves approximately $2NE_{trx}(1 - CH/N)/\tau$ of energy compared to the multihop, where N is the number of sensor nodes and CH is the number of cluster heads, and E_{trx} is the energy of the during transmission. Also, the latency of the clustered network is approximately equal to the time it takes a packet to move two hops. But, it increases with an increase in the path length for the multihop network. In the future, we shall add the plant growth model, fluid mechanics for the flowing nutrient solution, and nutrient consumption rate of different plants. We shall investigate how IoT and WSN could alleviate the contemporary hydroponics problems such as; maintenance cost, installation cost, plant disease detection, and plant disease prevention.

6 Derivation of a Model for the Proposed System

The instantaneous energy consumed ($E(t)$) by the sensor node is the sum of:

1. Energy consumed by sensors during sensing (E_s).
2. Energy consumed by sensors while sleeping (E_{sleep}), when sensor node is not sensing the environment.
3. Energy consumed by the sensor node during transmission of data (E_{tr}).
4. Energy consumed by the sensor node when forwarding data from N number of children (E_f).

Other parameters are:

1. Initial Battery Energy (E_{Init}).
2. Energy Consumed by the Sensor Nodes (E_{SN}).
3. Power consumed during sensing (P_{ss}).

$$E(t) = E_{Init} - E_{SN} \quad (17)$$

$$\text{Let } n = \text{number of times the sensor nodes sense the environment over a period } t \quad (18)$$

$$\begin{aligned} \Rightarrow E(t) &= E_{total} - \\ & (n(E_s + E_{tx} + N(E_{rx} + E_{tx}) + E_{sleep})) \\ \text{But } n &= \lfloor \frac{t}{\tau} \rfloor \\ \therefore E(t) &= E_{total} - \end{aligned} \quad (19)$$

$$\begin{aligned} & (\lfloor \frac{t}{\tau} \rfloor (E_s + E_{tx} + N(E_{rx} + E_{tx}) + E_{sleep})) \\ \text{where } E_s &= \sum_{i=1}^n E_i, \end{aligned} \quad (20)$$

$$\text{But, } E_{sleep} = P_{ss}t - nP_{ss}t_s \quad (21)$$

$$\Rightarrow E(t) = E_{total} - \lfloor \frac{t}{\tau} \rfloor (E_s + E_{tx} + \quad (22)$$

$$\begin{aligned} & N(E_{rx} + E_{tx})) - P_{ss}t - nP_{ss}t_s \\ \text{Let } E_{ss} &= P_{ss}t_s \end{aligned} \quad (23)$$

$$\therefore E(t) = E_{total} - \lfloor \frac{t}{\tau} \rfloor (E_s + E_{tx} + N(E_{rx} + \quad (23)$$

$$E_{tx})) - P_{ss}t - \lfloor \frac{t}{\tau} \rfloor E_{ss}$$

$$= E_{total} - \lfloor \frac{t}{\tau} \rfloor (E_s - E_{ss} + E_{tx} + \quad (24)$$

$$\begin{aligned} & N(E_{rx} + E_{tx})) - P_{ss}t \\ \text{if } E_{tx} &\approx E_{rx} \approx E_{trx}, \end{aligned} \quad (24)$$

$$\text{and } \lfloor \frac{t}{\tau} \rfloor \approx \frac{t}{\tau}$$

$$\text{Then } E(t) \approx E_{total} - \quad (25)$$

$$t \left(\frac{(E_s - E_{ss} + (1 + 2N)E_{trx})}{\tau} + P_{ss} \right) \quad (26)$$

$$\text{Let } \alpha = \frac{E_s}{E_{trx}}$$

$$\text{and } \beta = \frac{E_{ss}}{E_{trx}}$$

$$\begin{aligned} \text{Hence } E(t) &\approx E_{total} - \\ & t \left(\frac{(\alpha - \beta + (1 + 2N))}{\tau} E_{trx} + P_{ss} \right) \end{aligned} \quad (27)$$

$$\therefore \text{slope} = \frac{dE(t)}{dt} \quad (27)$$

$$\approx \frac{(E_s - E_{ss} + (1 + 2N)E_{trx})}{\tau} + P_{ss}$$

$$\approx \frac{(\alpha - \beta + (1 + 2N))}{\tau} E_{trx} + P_{ss}$$

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors would like to thank the Tertiary Education Trust Fund (TETFund), Yobe State University (YSU) Damaturu Nigeria and King Fahd University of Petroleum and Minerals (KFUPM) for their support in this research.

References

- [1] T. Baras, *DIY Hydroponic Gardens: How to Design and Build an Inexpensive System for Growing Plants in Water*, Cool Springs Press, 2018.
- [2] M. Stone, *How To Hydroponics: A Beginner's and Intermediate's In Depth Guide To Hydroponics*, Martha Stone, 2014.
- [3] Y. Nishikant, "Hydroponics Market by Type (Aggregate Systems, Liquid Systems), Crop Type (Vegetables, Fruits, Flowers), Equipment (HVAC, Led Grow Light, Irrigation Systems, Material Handling, Control Systems), Input, and Region – Global Forecast to 2025," 2019, accessed on 18th June, 2020.
- [4] S. Kumar, A. Sharma, S. Rana, "Vertical Farming: New Agricultural Approach for 21st Century," *Agri Mirror: Future India*, **1**(1), 27–68, 2020.
- [5] J. Muro, I. Irigoyen, P. Samitier, P. Mazuela, M. Salas, J. Soler, M. Urrestarazu, "Wood fiber as growing medium in hydroponic crop," in *ISSCH*, 179–185, Almeria, Spain, 2004, doi:10.17660/ActaHortic.2005.697.22.
- [6] T. Alexander, D. Parker, *The Best of Growing Edge*, New Moon Pub, 1994.
- [7] F. Aliyu, T. Sheltami, "Development of an energy-harvesting toxic and combustible gas sensor for oil and gas industries," *Sensors and Actuators B: Chemical*, **231**, 265 – 275, 2016, doi:10.1016/j.snb.2016.03.037.
- [8] J. Ko, C. Lu, M. B. Srivastava, J. A. Stankovic, A. Terzis, M. Welsh, "Wireless sensor networks for healthcare," *Proceedings of the IEEE*, **98**(11), 1947–1960, 2010, doi:10.1109/JPROC.2010.2065210.
- [9] M. A. Hussain, K. kyung Sup, et al., "WSN research activities for military application," in *ICACT*, 271–274, IEEE, Gangwon, Korea (South), 2009.
- [10] Y. Wang, X. Yin, D. You, "Application of wireless sensor networks in Smart Grid," *power System technology*, **34**(5), 7–11, 2010.
- [11] R. Hussain, J. Sahgal, P. Mishra, B. Sharma, "Application of WSN in rural development, Agriculture water management," *International Journal of Soft Computing and Engineering (IJSCE)*, **2**(5), 68–72, 2012.
- [12] U-Tec, "Green Farm," 2020, accessed on 18th June, 2020.
- [13] green-farm hydroponics, "Green Farm - Grow Your Own Herbs and Vegetables! – Organic fertilizers – do they work with hydroponic," 2015, accessed on 18th June, 2020.
- [14] U-ING, "Cube Green Farm Hydroponic Grow Box by U-ING," 2016, accessed on 18th June, 2020.
- [15] H. Twiggs, "The 9 best home hydroponics kits," 2018, accessed on 18th June, 2020.
- [16] Z. Eran, B. Elad, B. Roy, "Terraplanter," 2020, accessed on 18th June, 2020.
- [17] M. Saa'id, N. Yahya, M. Noor, M. M. Ali, "A development of an automatic microcontroller system for Deep Water Culture (DWC)," in *ICSPA*, 328–332, IEEE, Kuala Lumpur, Malaysia, 2013.
- [18] M. F. Saa'id, A. Sanuddin, M. Ali, M. S. A. I. M. Yassin, "Automated pH controller system for hydroponic cultivation," in *ISCAIE*, 186–190, IEEE, Langkawi, Malaysia, 2015, doi:10.1109/ISCAIE.2015.7298353.
- [19] F. Modu, A. Adam, F. Aliyu, A. Mabu, M. Musa, "A Survey of Smart Hydroponic Systems," *Advances in Science, Technology and Engineering Systems Journal*, **5**(1), 233–248, 2020.
- [20] DFRobot, "WATER TURBINE GENERATOR (5VDC)," 2017, accessed on 18th June, 2020.
- [21] Voltaic, "Voltaic Systems 1W 6V 113x89mm DRAWING CURRENT 2017 7 20," 2017, accessed on 18th June, 2020.
- [22] T. Asao, H. Kitazawa, K. Tomita, K. Suyama, H. Yamamoto, T. Hosoki, M. Pramanik, "Mitigation of cucumber autotoxicity in hydroponic culture using microbial strain," *Scientia Horticulturae*, **99**(3), 207 – 214, 2004, doi:10.1016/S0304-4238(03)00098-0.
- [23] W. Baudoin, R. Duffy, *Good Agricultural Practices for Greenhouse Vegetable Crops: Principles for Mediterranean Climate Areas*, Food and Agriculture Organization of the United Nations, 2013.
- [24] J. Jones, Chapter seven: Hydroponic Application factors, Taylor & Francis, 2014.
- [25] J. Jones, *Hydroponics: A Practical Guide for the Soilless Grower*, CRC Press, 2016.
- [26] G. L. Barbosa, F. D. A. Gadelha, N. Kublik, A. Proctor, et al., "Comparison of land, water, and energy requirements of lettuce grown using hydroponic vs. conventional agricultural methods," *International journal of environmental research and public health*, **12**(6), 6879–6891, 2015.
- [27] H.-J. K., W.-K. K., M.-Y. R., C.-I. K., et al., "Automated sensing of hydroponic macronutrients using a computer-controlled system with an array of ion-selective electrodes," *Computers and Electronics in Agriculture*, **93**, 46 – 54, 2013, doi:10.1016/j.compag.2013.01.011.
- [28] W.-J. C., H.-J. K., D.-H. J., D.-W. K., et al., "On-site ion monitoring system for precision hydroponic nutrient management," *Computers and Electronics in Agriculture*, **146**, 51 – 58, 2018, doi:10.1016/j.compag.2018.01.019.
- [29] T. Nishimura, Y. Okuyama, A. Satoh, "High-accuracy and low-cost sensor module for hydroponic culture system," in *GCCE*, 1–4, IEEE, Kyoto, Japan, 2016, doi:10.1109/GCCE.2016.7800514.
- [30] T. Nishimura, Y. Okuyama, A. Matsushita, H. Ikeda, A. Satoh, "A compact hardware design of a sensor module for hydroponics," in *GCCE*, 1–4, IEEE, Nagoya, Japan, 2017, doi:10.1109/GCCE.2017.8229255.
- [31] S. Charumathi, R. M. Kaviya, J. Kumariyarsi, R. Manisha, et al., "Optimization and Control of Hydroponics Agriculture using IOT," *Asian J. Appl. Sci. Technol*, **1**(2), 96–98, 2017.
- [32] R. Nalwade, T. Mote, "Hydroponics farming," in *ICEI*, 645–650, IEEE, Tirunelveli, India, 2017, doi:10.1109/ICOEI.2017.8300782.
- [33] S. Alam, S. T. Siddiqui, A. Ahmad, R. Ahmad, M. Shuaib, "Internet of Things (IoT) Enabling Technologies, Requirements, and Security Challenges," in *ADIS*, 119–126, Springer, Singapore, 2020, doi:10.1007/978-981-15-0694-9_12.
- [34] B. Siregar, S. Efendi, H. Pranoto, R. Ginting, U. Andayani, F. Fahmi, "Remote monitoring system for hydroponic planting media," in *ICISS*, 1–6, IEEE, Tangerang, Indonesia, 2017, doi:10.1109/ICTSS.2017.8288884.
- [35] Cooking Hacks, "Open Garden - Hydroponics & Garden Plants Monitoring for Arduino," 2014, accessed on 13th August, 2020.
- [36] Cooking Hacks, "Open Garden Hydroponics," 2014, accessed on 13th August, 2020.
- [37] Y.-B. Lin, Y.-W. Lin, C.-M. Huang, C.-Y. Chih, P. Lin, "IoTtalk: A management platform for reconfigurable sensor devices," *IEEE Internet of Things Journal*, **4**(5), 1552–1562, 2017, doi:10.1109/JIOT.2017.2682100.
- [38] T.-H. Wu, C.-H. Chang, Y.-W. Lin, L.-D. Van, Y.-B. Lin, "Intelligent plant care hydroponic box using IoTtalk," in *GreenCom*, 398–401, IEEE, Chengdu, China, 2016, doi:10.1109/iThings-GreenCom-CPSCoM-SmartData.2016.94.
- [39] L. Van, Y. Lin, T. Wu, Y. Lin, et al., "PlantTalk: A Smartphone-Based Intelligent Hydroponic Plant Box," *Sensors*, **19**(8), 1763, 2019, doi:10.3390/s19081763.
- [40] K. E. Lakshmiprabha, C. Govindaraju, "Hydroponic-based smart irrigation system using Internet of Things," *International Journal of Communication Systems*, **1**(1), 1–10, 2019, doi:10.1002/dac.4071.
- [41] M. A. Maureira, D. Oldenhof, L. Teernstra, "ThingSpeak—an API and Web Service for the Internet of Things," 2011, accessed on 2nd September, 2020.
- [42] P. N. Crisnapati, I. N. K. Wardana, I. K. A. A. Aryanto, A. Hermawan, "Hommons: Hydroponic management and monitoring system for an IOT based NFT farm using web technology," in *CITSM*, 1–6, IEEE, Denpasar, Indonesia, 2017, doi:10.1109/CITSM.2017.8089268.
- [43] M. A. Triawan, H. Hindersah, D. Yolanda, F. Hadiatna, "Internet of things using publish and subscribe method cloud-based application to NFT-based hydroponic system," in *ICSET*, 98–104, IEEE, Bandung, Indonesia, 2016, doi:10.1109/ICSEngT.2016.7849631.
- [44] S. Ruengittinun, S. Phongsamsuan, P. Sureeratanakorn, "Applied internet of thing for smart hydroponic farming ecosystem (HFE)," in *Ubi-Media*, 1–4, IEEE, Pattaya, Thailand, 2017, doi:10.1109/UMEDIA.2017.8074148.
- [45] A. Satoh, "A Hydroponic Planter System to enable an Urban Agriculture Service Industry," in *GCCE*, 281–284, IEEE, Nara, Japan, 2018, doi:10.1109/GCCE.2018.8574661.
- [46] J. A. Manrique, J. S. Rueda-Rueda, J. M. T. Portocarrero, "Contrasting Internet of Things and Wireless Sensor Network from a Conceptual Overview," in *iThings*, 252–257, IEEE, Chengdu, China, 2016, doi:10.1109/iThings-GreenCom-CPSCoM-SmartData.2016.66.
- [47] Philips, "Get started," Accessed on 27th February, 2022.

- [48] M. Zennaro, B. Pehrson, A. Bagula, "Wireless Sensor Networks: a great opportunity for researchers in Developing Countries," in WCITD, 1–7, WCITD, Pretoria, South Africa, 2008.
- [49] S. Farahani, ZigBee wireless networks and transceivers, Newnes, 2011.
- [50] R. Prabha, R. S. Saranish, S. Sowndharya, A. Santhosh, R. Varsha, K. Sumathi, "IoT Controlled Aquaponic System," in ICACCS, 376–379, IEEE, Coimbatore, India, 2020, doi:10.1109/ICACCS48705.2020.9074401.
- [51] S. Diver, L. Rinehart, "Aquaponics-Integration of hydroponics with aquaculture," 2000, accessed on 18th June, 2020.
- [52] S. V. Souza, R. M. T. Gimenes, E. Binotto, "Economic viability for deploying hydroponic system in emerging countries: A differentiated risk adjustment proposal," Land Use Policy, **83**, 357 – 369, 2019, doi:10.1016/j.landusepol.2019.02.020.
- [53] M. Kocakulak, I. Butun, "An overview of Wireless Sensor Networks towards internet of things," in CCWC, 1–6, IEEE, Las Vegas, NV, USA, 2017, doi: 10.1109/CCWC.2017.7868374.
- [54] M. Martinez-Mate, B. Martin-Gorriz, V. Martínez-Alvarez, M. Soto-García, et al., "Hydroponic system and desalinated seawater as an alternative farm-productive proposal in water scarcity areas: Energy and greenhouse gas emissions analysis of lettuce production in southeast Spain," Journal of Cleaner Production, **172**, 1298 – 1310, 2018, doi:10.1016/j.jclepro.2017.10.275.
- [55] S. Tagle, R. Pena, F. Oblea, H. Benzoza, et al., "Development of an Automated Data Acquisition System for Hydroponic Farming," in HNICEM, 1–5, Baguio City, Philippines, 2018, doi:10.1109/HNICEM.2018.8666373.
- [56] D-Robotics UK, "DHT11 Humidity & Temperature Sensor," Datasheet, 1–9, 2010.
- [57] S. A. D., K. F., J. R., R. F., et al., "A survey on data aggregation techniques in IoT sensor networks," Wireless Networks, **26**(2), 1243–1263, 2020, doi: 10.1007/s11276-019-02142-z.
- [58] Omnet.org, "What is OMNET++," 2020, accessed on 2nd September, 2020.
- [59] F. M. Aliyu, F. Modu, A. S. Adam, A. Mabu, M. A. Musa, "Simponics," 2020, accessed on 2nd September, 2020.
- [60] E. C., C. Jose, G. C., "Modeling of current consumption in 802.15. 4/ZigBee sensor motes," Sensors, **10**(6), 5443–5468, 2010, doi:10.3390/s100605443.
- [61] Atlas Scientific, "EZO-ECTM Embedded Conductivity Circuit," Atlas Scientific LLC, **6**(1), 1–75, 2019.
- [62] Atlas Scientific, "Conductivity Probe K 10," Atlas Scientific LLC, **2**(7), 1–13, 2020.
- [63] Parallax Inc, "Specification for 1W PV Module," 2020, accessed on 2nd September, 2020.
- [64] Atlas Scientific, "Micro Footprint pH Monitoring Subsystem," Atlas Scientific LLC, **4**(0), 1–13, 2020.
- [65] Atlas Scientific, "Lab Grade pH Probe Double Junction Silver/Silver Chloride with EXR Glass," Atlas Scientific LLC, **4**(2), 1–13, 2020.
- [66] V. Dale, "NTC Thermistors, Resistance/Temperature Conversion, Curve 2," 2010, accessed on 2nd September, 2020.
- [67] D. W. Medugu, F. W. Burari, A. A. Abdulazeez, "Construction of a reliable model pyranometer for irradiance measurements," African journal of Biotechnology, **9**(12), 2010.

Enhanced Dynamic Cross Layer Mechanism for real time HEVC Streaming over Vehicular Ad-hoc Networks (VANETs)

Marzouk Hassan*, Abdelmajid Badri, Aicha Sahel, Belbachir Kochairi, Nacer Baghdad

Electrical Engineering Department, Faculty of Science and Technology, Hassan II University, Mohammedia City, Morocco

ARTICLE INFO

Article history:

Received: 30 November, 2021

Accepted: 26 January, 2022

Online: 09 March, 2022

Keywords:

Ross layer

Video transmission

PSNR

VANET

Video transmission

ABSTRACT

Various applications have helped make vehicular Ad-hoc network communication a reality. Real-time applications, for example, need broadcasting in high video quality with minimal latency. The new High-Efficiency Video Coding (HEVC) has shown great promise for real-time video transmission through Vehicle Ad-hoc Networks due to its high compression level. These networks, on the other hand, have highly changeable channel quality metrics and limited capacity, making it challenging to maintain good video quality. HEVC real-time video streaming on VANET may now benefit from an end-to-end dynamic adaptive cross-layer method. According to the video coding process's time prediction structure, frame size, and network density, each video packet should be assigned to a suitable Access Category (AC) queue on the Medium Access Control layer (MAC). The results we've gotten demonstrate that the new method suggested delivers considerable improvements in video quality at end-to-end latency and reception in comparison to the Enhanced Distributed Channel Access (EDCA) specified in the 802.11p standard for several targeted situations. Quality of Experience (QoE) and Quality of Service (QoS) assessments have been used to verify our proposed strategy.

1. Introduction

As the idea of a city linked to the internet becomes closer to reality, the effect of the internet on our lives grows. Nowadays this may be realized with the appropriate use of traffic safety and entertainment applications in the form of vehicular networks. Inter-vehicle or infrastructure communication network may be used for a variety of purposes, but one of the most intriguing is video streaming. For this reason, it isn't easy to broadcast video through automobile networks. The transmission of video content over vehicle networks would represent a big step forward [1]; Overtaking maneuvers, parking assistance, video communication, video surveillance, and public transport assistance, and for entertainment, the possibility to use visual information data [2], [3]. However, compressed videos are susceptible to noise and channel loss. Although virtual networks are plagued by harsh transmission circumstances and packet loss rates (PLR) that do not ensure the quality of service, there are other issues.

Several technological solutions have been suggested to improve multimedia transmissions over vehicle networks [4]. Particularly, the IEEE 802.11p standard, which has been solely dedicated to vehicle networks, At the MAC layer, the standard

handles QoS differences by offering distinct service classes [5]. In contrast, the HEVC/H265 standard has recently been developed and put at the disposal of scientists; this new standard outperforms its predecessor (H264/AVC) coding efficiency-wise by about 50% [6]. Due to the requirements of video transmission, inter-vehicle applications using video, like traffic optimization and monitoring, ensuring low delay has become essential [7], [8].

It is even more important in remote vehicle control applications and driver assistance systems [9], given the recent interest in autonomous vehicles. Therefore, a communication system ought to ensure both low latency and high reliability [10].

In a vehicle environment, the received signal intensity can vary considerably because of several factors; fading, shading, multipath, and Doppler effect are the main ones. Therefore, VANETs are networks with difficult channel conditions resulting in a degradation of the output of the link, which results in poor quality of the video. To address this, many studies have evaluated video quality as a network load function [11] or the video source encoder [12]. Authors in [13] suggested real-time performance assessment of video transmission in-vehicle environments. Specifically, their research looked at vehicle density and distance effects on HEVC-encoded video sequences in the road and urban environments. As assessment measures, the peak signal to noise

*Corresponding Author: Marzouk Hassan, marzouk.hsn@gmail.com

ratio (PSNR) and the packet delivery ratio (PDR) were calculated. A change to the Real-Time Transport Protocol (RTP) was developed by authors in [1] to make the H.264 encoded video transmission more efficient to enhance the transfer of information. The implementation of video transmission in VANET was also studied. Using a retransmission technique in [14] devised an error recovery mechanism. MPEG4 part 2 video is encoded with uneven protection of video images, according to the standard. Regarding video streaming through VANET networks, researchers in [15] employed network coding and blanking coding.

Improvements were also made to EDCA for video transmission on the IEEE 802.11e standard. Background traffic (BK), best effort (BE), which EDCA makes accessible in accordance with the meaning of video coding, were initially proposed by authors in [14] and have since been widely used. A mapping algorithm based on the IEEE 802.11e EDCA traffic standard was suggested by the authors to increase H.264 video transmission over an IEEE 802.11e network. But since this used mapping algorithm is static, it does not reflect the network state. IEEE 802.11e wireless networks might benefit from a dynamic cross-layer mapping technique developed in [16], which they believe would be effective. Authors in [17] created a cross-layer framework enabling H.264/AVC video streaming through IEEE 802.11e wireless networks, which was published in IEEE Communications Magazine. The suggested technique provides for more effective use of the radio source by assessing the access time for each AC and selecting the AC with the shortest access time. However, the work stated for cross-layer approaches is particular to the IEEE 802.11e standard and is grounded on the previous standards for video encoding; the video encoder's ability to cause modifications in the temporal standards prediction framework has not been taken into account. On top of that, they do not take into consideration the issue of latency for a low-delay transmission. Researchers in [18] developed a framework of delay rate distortion in wireless video communication employing H. 264's LD mode, which is constructed from predicted and intra frames, called P and I frames respectively. A real-time H.265/HEVC stream transmission technique was suggested by authors in [19]. The optimal time prediction is chosen by algorithms to be used by considering the decoding and encoding times of the Network QoS and HEVCs.

To optimize HEVC video streaming on VANETs, we have created a dynamic cross-layer technique. We propose a mapping mechanism that is devoted to the IEEE 802.11p standard to increase the efficiency of video streaming in Vehicular Adhoc networks with fluctuating network topology. HEVC's new temporal prediction structures allow us to make use of our approach. The IEEE 802.11p and HEVC standards have influenced the re-design of the method initially described in [20] and [16]. Both the relevance of the channel state and the video frame, controlled by the queueing system of the MAC layer, are taken into account by the suggested approach. Taking into consideration the video's temporal prediction structure, frame significance, and current traffic load, each packet of the transmitted video is assigned to the most suitable AC queue on the MAC layer.

Section 2 highlights our proposed solution in detail. In section 3, we will focus on our work approach and simulation. Section 4 contains the simulation results that demonstrated the proposed solution's effectiveness, providing 18% average received packet

gain in comparison to the IEEE 802.11p EDCA mechanism. Conclusion is described in the last section (Section 5).

2. Description of the proposed solution

For the purpose of achieving considerable performance advantages, cross-layer design refers to a method that takes advantage of the reliance across protocol levels. Depending on how information is shared across layers, several different design types may be identified. Authors in [21] narrowed the range of feasible designs down to four distinct methods. Using the first way, new interfaces are created. The second involves merging nearby layers, the third consists of the designed integrating with new interfaces, and the last approach involves the vertical calibration across the layers.

The proposed cross-layer architecture takes use of information about video packets' relevance obtained from the application layer to regulate this on the decision-making process at the MAC layer when video packets are considered necessary. The technologies that were used in this project will be discussed in further detail later in this section. We will begin by discussing the properties of IEEE 802.11p, which are unique to vehicle networks, and then move on to more general considerations. As a second step, we will offer a high-level overview of H.265/HEVC encoding before presenting our suggested cross-layer architecture.

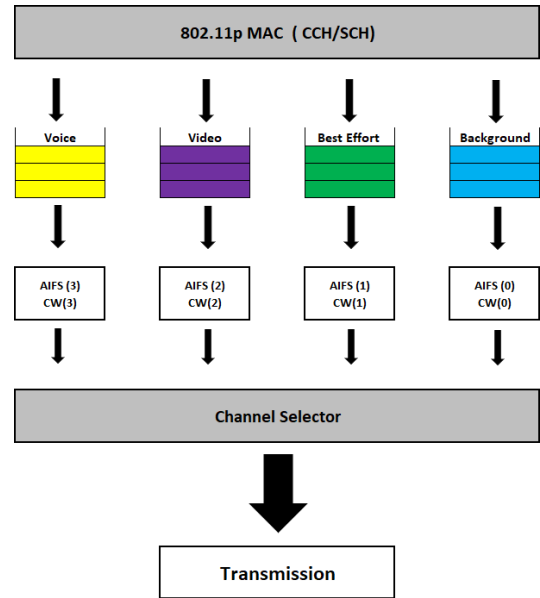


Figure 1: The different access categories in the IEEE802.11p MAC architecture.

2.1. The IEEE 802.11p standard

The IEEE 802.11p standard is an accepted addition to the IEEE 802.11 standard for providing wireless connectivity in a vehicle context. It was approved by the IEEE in 2009. (WAVE). The standard's PHY layer is based on the DSRC (dedicated short-range communication) standard. It operates in the 5.850-5.925 GHz frequency band, with a IEEE 802.11a modified version serving as the physical layer. According to [22], DSRC is regarded to be capable of providing communication for both vehicular to infrastructure (V2I) and vehicle to vehicular (V2V) situations. The European Standard Telecommunications Institute (ETSI) describes ITS-G5 as the comparable standard in Europe to the

IEEE 802. p standard, which is devoted to the United States [23]. There are some discrepancies between the two standards at the higher levels, although they are minor. Despite this, it operates in the same frequency range as the DSRC [24]. In Japan, the equivalent of the DSRC is utilized in the 5.8 GHz frequency band, which is composed of six service channels (SCH) and one control channel. It also uses a 3 Mbps preamble supports data speeds of 3, 6, 9, 12, 18, 24, and 27 megabits per second. Orthogonal Frequency Division Multiplexing (OFDM) is the modulation technique used (OFDM).

The IEEE 802.11p standard's medium access control layer protocol employs CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) as the principal medium access mechanism for link sharing and EDCA for packet transport [25]. The EDCA protocol, in conjunction with flow prioritizing in accordance with QoS criteria [26], facilitates service hierarchization. The IEEE.802.11e standard was first introduced, and it has since undergone several revisions [27][20]. Actually, EDCA is an advance over the distributed channel access (DCA) technique to provide the necessary quality of service (QoS). A single queue for holding data frames is replaced by four queues, each indicating a distinct degree of priority or access category, referred to as ACs in this document. Every one of these acs is allocated to a certain kind of traffic, as depicted in Fig.1, with the background (BK), video (VI), voice (VO), and best effort (BE) being examples.

The higher the transmission priority, the greater the likelihood of successful transmission. Priority is allocated to each traffic stream by the relevance of that traffic stream. Priority has been given to VoIP traffic, which was followed by video, background traffic, and best-effort, all of which had lower priority.

The waiting time TAIFS (Time Arbitration Inter-Frame Space), which represents the time required for each AC to access the media, is used to determine the priority of each AC. It enables varying prioritizing of frames based on the kind of traffic being sent. Time between frames may be reduced by using a short TAIFS, for example, and the time required to connect to the medium. TAIFS value is given by [27]:

$$T A I F S [A C] = A I F S N [A C] * a S l o t T i m e + S I F S \quad (1)$$

The AIFSN [AC] (Arbitration Inter-Frame Space Number) is the constant that corresponds to each AC, which is the AC of each traffic type. There are specified consistent intervals for the Short Inter-Frame Space and aSlotTime in the standard, 32 and 13 second time frames. The contention windows are another distinction between the ACs (CW).

Internal queue clashes are possible since EDCA has four queues. The process mentioned before aids in the resolution of these issues. Figure 2 displays an illustration of competing for access to the media and the TAIFS prioritizing system. As can be seen, a best effort frame and a voice frame are in a heated competition for access to the media. In order to reach the medium, the AC voice's reduced wait time allows it to forego its best effort. Each AC's value is listed in Table 1[27]. Distinct ACs have different CW and AIFSN values set in the CCH and SCH. According to smaller TAIFS, we conclude that video AC has a higher priority than the BK and BE.

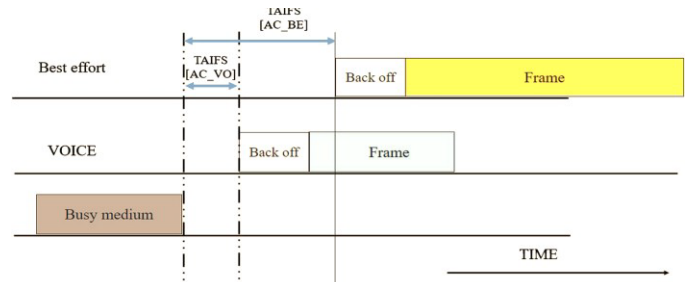


Figure 2: the medium access in EDCA IEEE802.11p.

2.2. Encoding modes for H265/HEVC

HEVC, like its predecessor H264/AVC, follows a hybrid video coding scheme. Both video coding standards have a two-layered high-level design consisting of a network abstraction layer (NAL) and video coding layer (VCL). The VCL includes all low-level signal processing, including inter-and intra-picture prediction, block partitioning, transform coding, in-loop filtering, and entropy coding. At the top-level, an HEVC sequence consists in a series of network adaptation layer (NAL) Units or NALUs. These NALUs encapsulate compressed payload data and include parameter sets containing key parameters used by the decoder to correctly decode the video data slices, which are coded video frames or parts of video frames [16].

It is conceivable to envision video transmission, especially in real-time, in networks with little capacity or a high packet loss rate because of the general benefit of HEVC. As it is considered a hostile network, strong level of resistance and compression is required for transmission of video in the VANET. This is because transmission of video in the VANET is considered to be pretty hostile. HEVC has been shown by researchers in [19] to exceed its predecessors significantly when it comes to decreasing temporal error propagation in changeable wireless video environments. Their study compared the HEVC encoding pattern with an LD configuration to the traditional MPEG-4 part 2, H.264/AVC, and H.263 coding standards under various packet loss rates.

Predictions from future pictures are prohibited to ensure low latency operations at both the decoder and encoder. While the short-latency restriction may be met by employing P-images solely, the directional motion compression efficiency estimate is lost due to this practice. Generalized P-B (GPB) pictures are introduced in HEVC to reduce the time to process a B- picture while still delivering excellent coding performance [25]. A GPB is a bi-predictive frame that employs just previous pictures for inter-prediction in GPBs.

Error-resilience, processing time, computational complexity, codec efficiency, and approaches are all considered while configuring HEVC for a specific application. The two most common encoding setups are:

- the “high efficiency” approach that provides highly efficient coding with a significant computational cost,
- Excellent efficiency with little coder complexity in the “low complexity” mode.

2.3. Proposed cross-layer approach description

Our multilayer system is described in this section. Video transmission at the MAC layer of the IEEE 802.11p standard is limited to the use of the specialized video AC. The other two lower

priority ACs may be used to reduce network congestion and the loss of video packets due to video packet overflow.

When it comes to our system’s current development, we are still working on the low latency element. To do this, we’re looking at two low-complexity video transmission techniques:

- Static inter-layer mapping algorithm that is centered on hierarchical HEVCencoding. (Figure 3)
- Adaptive inter-layer mapping algorithm that is developed based on hierarchicalHEVC encoding. (Figure 4)

The Cross-layer system also uses the HEVC hierarchy to map video packets at the IEEE 802.11p standard MAC layer. It is demonstrated that the three levels of stratification in the two suggested multilayer mapping methods are based on the video structure:

For the low delay configuration

- Layer-1: includes level 0 images and level I images
- Layer-2: includes level 1 executives.
- Layer-3: includes level 2 executives.

For the random-access configuration:

- Layer-1: comprises level 1 and level 0 images and I images.
- Layer-2: comprises level 2 as well as level 3 frames.
- Layer-3: comprises level 4 executives.

The choice of the distribution of the frames was established according to the importance, and the size of the frames compressed data. No categorization has been kept for the All Intra configuration.

a. Static mapping algorithm

According to the categorization system used, which changes based on the video structure, the pictures associated with layer-1 are the essential images. This is due to the fact that layer-1 pictures have a significant effect and, in some ways, influence everything else in GoP. In this sense, any loss or deterioration that may occur due to their actions will impact the whole GoP. Additionally, it's worth noticing that the Layer 1 photos include additional information. We recommend creating a static technique for each video structure based on this information. It is always assigned the highest priority for layer 1 frames to utilize alternating current, whereas layer 2 frames are always assigned the lowest priority. It's a video that's been made by AC. Route the second most critical Layer 2 frames to the second available queue, which is likely to have the best AC effort available. In this section, we will, however, stratify the video using the method proposed in [20]. When using the static method, we'll put video packets corresponding to layer 3 in the final queue (BAC).

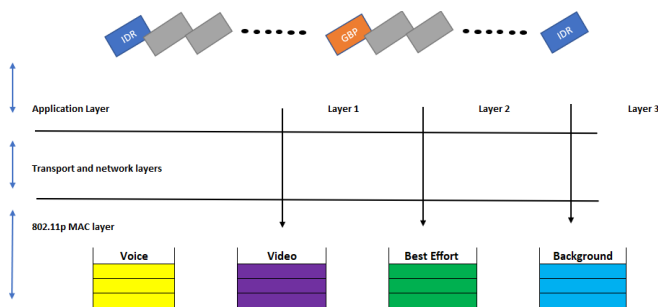


Figure 3: Illustration of the static cross-layer algorithm.

b. Adaptive mapping algorithm

Video packets are assigned the most suitable AC currents based on the suggested adaptive mapping method at the MAC layer of the network. Network traffic load, the relevance of each frame, and temporal prediction structure are all considered. As the last step, we must assign each picture type a separate mappings probability to lower priority ACs, denoted as P Layer. The probability is a function of the frame size meaning:

$$0 \leq P_{\text{Layer-1}} \leq P_{\text{Layer-2}} \leq P_{\text{Layer-3}} \leq 1$$

Alternatively, as previously stated, the channel’s condition affects the mapping. AC queues are a good indicator of network traffic congestion. To avoid overcrowding, it is essential to keep the MAC queue buffer as empty as possible. Random Early Detection (RED) is the philosophy behind the two thresholds that we’ve implemented to manage and minimize network congestion. According to [20], the adaptive mapping method is based on the following formula:

$$P_{\text{new}} = P_{\text{Layer}} \times \frac{qlen(AC [VI]) - qthlow}{qthhigh - qthlow}$$

Qlen (AC [VI]) is the real length of the video queue, and qthlow and qthhigh, which are arbitrarily set thresholds, explicitly state the process and the degree of mapping of ACs of lower priority.

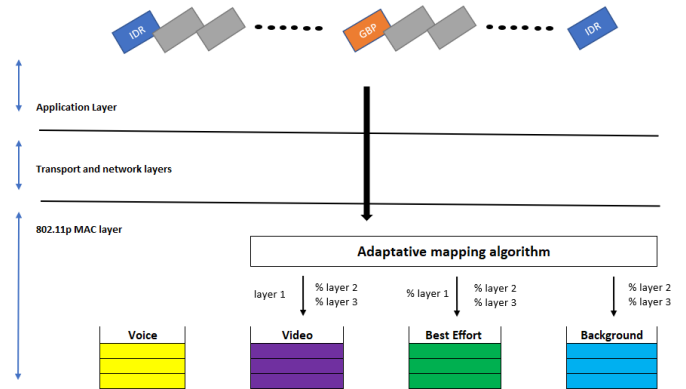


Figure 4: Illustration of the Adaptive mapping algorithm.

3. Framework and Simulation set-up

Integrating a map into a network simulator has been and will always remain a challenging task for researchers. OpenStreetMap was created by academics to tackle this problem and be used in traffic simulations. It is a free customizable map of the globe, has an incredible quantity of data, as well as a high degree of precision. However, since the data is frequently incomplete for traffic simulations, Map acquisition should always be the initial phase; followed by filling the missing sections and enhancing the data before turning it into an OSM file that the SUMO traffic simulator can use.

Figure 5 illustrates the four essential phases of our working method. The following section will discuss each stage in detail.

It is necessary to first download and install MPEG, which is an accepted practice for video streaming over the internet. For this simulation we have used a CIF (H.261) video file format with a 352 x 288. before a CIF file can be used for simulation, a video trace file is generated by running the mp4trace utility on the

original MPEG4 movie. If the picture has been segmented, the video trace file provides information on the segments' number, kind, and size. The mp4trace tool requires the port number and target URL since the Evalvid utility was initially developed to analyze real video transmissions.

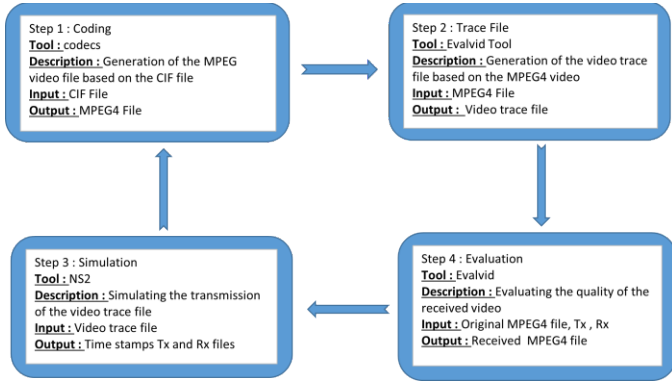


Figure 5: work approach

4. Simulation results

Different routing protocols will be tested in this simulation to see how they work when running in a high-density traffic environment. After 200 milliseconds of simulation, there will be ten to 60 cars, and on each simulation ten vehicles will be added to the total. For the simulation and as mentioned below, UDP was used as a transport layer protocol, and CBR as the application layer protocol.

Table 1: Simulation parameters of routing protocols performance evaluation

Parameters	
Simulator	NS-2.35
Protocols	AODV, DSDV, DSR, OLSR
Simulation duration	200s
Simulation area	3511m*3009m
Number of vehicles	10,20,30,40,50,60
MAC layer protocol	IEEE 802.11
Application layer protocol	UDP
Paquets size	CBR

And Casablanca's Anfa area was chosen for our simulation Figure 6.

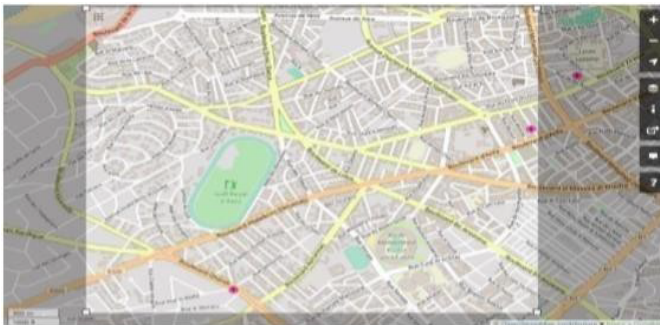


Figure 6: Anfa District Casablanca "OpenStreetMap"

In comparison to Destination Sequenced Distance Vector (DSDV) and Optimized Link State routing protocols (OLSR), Dynamic Source Routing (DSR) and On-demand Distance Vector routing protocols (AODV) have performed better, which is reasonable considering that proactive protocols must sustain a forwarding table for every node in the network. Through the VANETs high mobility, a large number of updates to the routing table must be made momentarily, resulting in bandwidth wastage.

Since it was important to see how well PSNR performed when streaming low-brightness videos, in the second phase of the simulation, we chose to proceed with the AODV protocol as it was the most efficient in terms of throughput, jitter, and packet delivery.

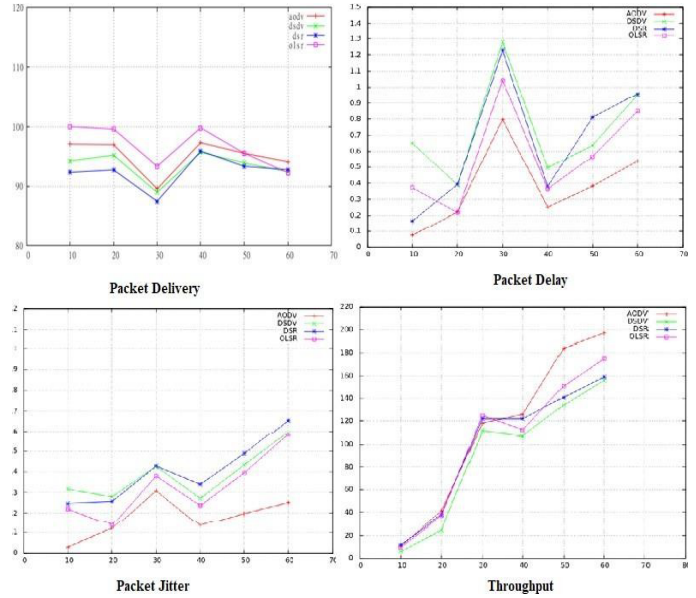


Figure 7: throughput, packet delay, packet delivery, and jitter, and for various network densities (5, 10,20,30,40, 50, 60)

Since it was important to see how well PSNR performed when streaming low-brightness videos, in the second phase of the simulation, we chose to proceed with the AODV protocol as it was the most efficient in terms of throughput, jitter, and packet delivery.

The other parameters are mentioned below:

Table 2: Simulation parameters of PSNR performance evaluation

Parameters	
MAC layer protocol	802.11
Routing protocol	AODV
Number of vehicules	4, 9, 25, 64
Image Resolution	352 * 288
Video file frame size	30 fps

"Highway CIF" is the video we utilized for our scenario. When the network sparsity is adjusted to $D = 100$ m and its density increases, the PSNR performance of the AOADV protocol is shown in Figure 8. To get a better picture of the data, we utilized 100 frames to smooth it down a little.

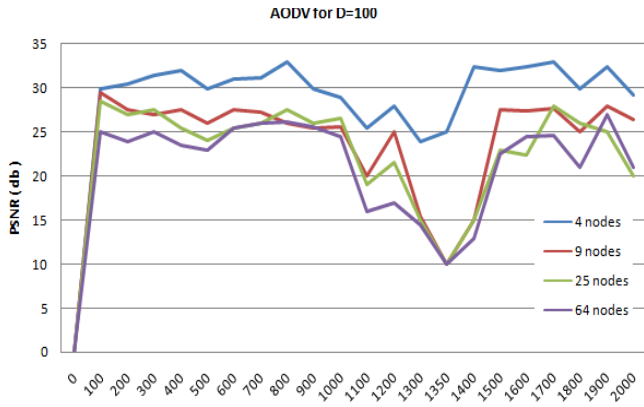


Figure 8: PSNR Performance of AODV for various network densities (4, 9, 25, 64)

At various densities of networks, the brightness of video pictures has an impact on maintaining the PSNR variation model. All frames in this movie have the similar brightness since we send the same video file over many topologies. As shown in Figure 8, this was confirmed by looking at the two significant decreases in PSNR performance. The first happened at a frame rate of around $F = 500$, or about 19 seconds into the movie viewing. As shown in Figure 9, the brightness reduced during this time due to the emergence of a black automobile in excess. During video playing times of $T = 43$ and $T = 46$, the second big reduction in PSNR occurred. In the video, a black bridge initially emerges. After that, the automobile passes over its shadow, as seen in Figure 6. PSNR falls in both circumstances because as a frame’s brightness content drops, noise energy outweighs maximum signal energy, so lower PSNR may be attributed to this fact.



Figure 9: screenshot of the video at $T=21s$ and $T=41s$

Figure 8 shows that the performance of PSNR of the AODV decreases with the increase in network density. When the network grows from $N = 4$ to $N = 9$, the PSNR drops by around 5 dB between Frame = 650 and Frame = 500. However, as the number of nodes in the network changes from $N = 25$ to $N = 64$, this attenuation is less relevant. Data must be routed through intermediary nodes when the network density rises to $N = 9, 25$, or 64 nodes. In this scenario, the PSNR suffers greatly because of the many jumps.

To conclude our simulations, we performed several tests to show the suggested mechanism’s efficacy. Fig.10 illustrates the benefit of the adaptive method. The PSNR curves show how the two mapping techniques change over time. In terms of PSNR, the adaptive technique (red) is superior in performance. In certain peaks, the static approach yields strong PSNR values, indicating good receipt of IDR pictures. This, However, doesn’t apply to the remaining GoP frames which have bad PSNR score. On the other hand, the video quality is still superior to that of the EDCA approach. In addition to the latter, a GoP’s intra-frame reference picture loss may be seen in Fig.10. When the initial frame of a GoP is lost, the PSNR of the whole GoP is reduced. To further

investigate this point, we have used a portion of the graph to see the states we’ve already examined.

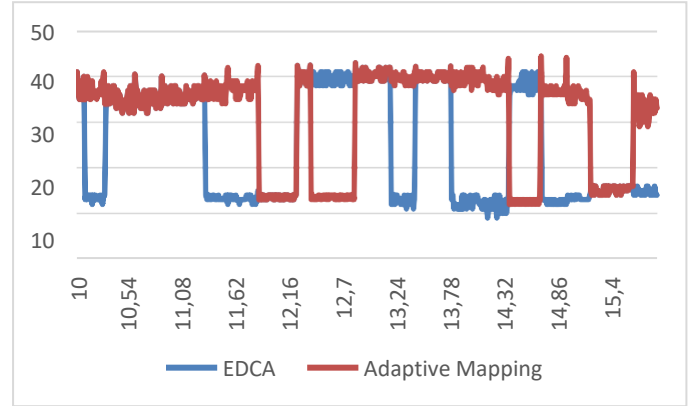


Figure 10: The variation of the PSNR for the different mapping algorithms: EDCA (blue) and adaptive (red)

Table 3: Average PSNR and number of packets lost for each mapping algorithm.

Mapping algorithm	Average PSNR	Number of lost packets			
		Layer-1frame	Layer-2frame	Layer-3frame	Total
EDCA	23.86	19	19	37	75
Adaptative mapping	31.71	2	3	6	11

Due to the classification of video packets, and the usage of IEEE 802.11p standard resources, packet losses may be minimized, and the most critical video packets can be protected. For example, if a technique is more efficient, the overall number of lost packets reduces dramatically. The adaptive technique has a packet loss rate of 11 compared to 75 for the EDCA method. The EDCA’s packet loss is evenly distributed throughout the several tiers. The unbalance also depends on the relevance of the layer for both static and adaptive approaches. There are just two missed packets when using the adaptive method instead of the EDCA’s 19 when using the static approach. The adaptive technique, on the other hand, can better secure the most critical layers’ packets ensuring a better video quality as seen by the average PSNR of a video sequence.

5. Conclusion

In this paper, we have investigated the combined effect of the network density and the image blitheness on the PSNR performance. We created a variety of network models with varied network densities, and the assessment results revealed several intriguing facts, including the fact that PSNR performance degrades as the network density grows. It is also discovered that the PSNR suffers a significant reduction when the network density rises due to packet loss.

Video transmission in a vehicular environment is affected by various forms of losses, which results in packet loss and greatly affects the perception of perceived quality. The real-time transmission of a live video feed via the VANET is a difficult task. However, the new HEVC coder shows more promising results and offers considerable advancements in video coding in a wireless

setting compared to its predecessor. Adaptive algorithms are presented in this study.

Low-latency HEVC streaming over IEEE 802.11p vehicular networks may now be improved using a new cross-layer mapping approach. MAC layer application layer information is used in a cross-layer manner in the suggested enhancement. Indeed, the method can optimally transport video packets based on information about the MAC layer buffer filling status, frame type, and temporal prediction video structure.

Simulation findings reveal that the suggested alternatives outperform the typical EDCA in many distinct situations and scenarios. In addition, a comparison of the suggested adaptive algorithm's QoS and QoE results showed that it gives the best outcomes for the various HEVC temporal forecast structures.

The present AI encoding setup does not include any kind of categorization. As a result, our next step would be to look into a more efficient video packet classification algorithm for this kind of transmission. Also, packets that aren't received in the allocated time aren't included in the calculation. As a result, sending them through the network is a waste of time and bandwidth, therefore they can be eliminated at the transmitter. Hence an algorithm capable of doing so should be considered, an algorithm connects the queue buffering time, delay constraints at application level and end to end delay.

References

- [1] M.G. W.L. Junior, D. Rosário, E. Cerqueira, L.A. Villas, "A game theory approach for platoon-based driving for multimedia transmission in VANETs," *Wirel. Commun. Mob. Comput.*, **2414658**, 1–11, 2018, doi:https://doi.org/10.1155/2018/2414658.
- [2] A.V.V. M. Jiau, S. Huang, J. Hwang, "Multimedia services in cloud-based vehicular networks," *IEEE Intell. Transport. Syst. Mag.*, **7**(3), 62–79, 2015, doi:https://doi.org/10.1109/MITS.2015.2417974.
- [3] X.Z. M. Gerla, C. Wu, G. Pau, "Content distribution in VANETs, Veh.," (*Veh. Commun.* **1**(1), 3–12, 2014.
- [4] R.S. C. Campolo, A. Molinaro, "From today's VANETs to tomorrow's planning and the bets for the day after," (*Veh. Commun.* **2**(3), 158–171, 2015, doi:https://doi.org/10.1016/j.vehcom.2015.06.002.
- [5] R.F.S. D. Perdana, "Performance comparison of IEEE 1609.4/802.11p and 802.11e with EDCA implementation in MAC sublayer," in *International Conference on Information Technology and Electrical Engineering (ICITEE)*, 285–290, 2013.
- [6] T.W. G.J. Sullivan, J.R. Ohm, W.J. Han, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.* **22**(12), 1649–1668, 2012, doi:https://doi.org/10.1109/TCSVT.2012.2221191.
- [7] H.D. I. Parvez, A. Rahmati, I. Guvenc, A.I. Sarwat, "A survey on low latency to-wards 5G: RAN RAN, core network and caching solutions," *IEEE Commun. Surv. Tutor.* (1), 2018, doi:https://doi.org/10.1109/COMST.2018.2841349.
- [8] E.C. C. Quadros, A. Santos, M. Gerla, "QoE-driven dissemination of real-time videos over vehicular networks," *Computer Communications*, **91–92**, 91–92, 2016.
- [9] M.F. P. Gomes, C. Olaverri-Monreal, "Making vehicles transparent through V2V video streaming," *IEEE Trans. Intell. Transp. Syst.* **13**(2), 930–938, 2012, doi:https://doi.org/10.1109/TITS.2012.2188289.
- [10] T.H. R. Alieiev, A. Kwoczek, "Automotive requirements for future mobile networks," *IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*, 1–4, 2015.
- [11] J.I.A. M. Oche, R.M. Noor, "Network centric QoS performance evaluation of IPTV transmission quality over vanets," *Comput. Commun.* **61**, 34–47, 2015, doi:https://doi.org/10.1016/j.comcom.2014.12.001.
- [12] M.P.M. P. Pinol, A. Torres, O. Lopez, M. Martinez, "Evaluating HEVC video delivery in VANET scenarios," *IFIP Wireless Days (WD)*, (Nov. 2013), 2013.
- [13] Y.J. A. Torres, C.T. Calafate, J.-C. Cano, P. Manzoni, "Evaluation of flooding schemes for real-time video transmission in VANETs."
- [14] D.A. I. Zaimi, Z.S. Houssaini, A. Boushaba, M. Oumsis, "An evaluation of routing protocols for vehicular ad-hoc network considering the video stream," *Wirel. Pers. Commun.*, **98**(1), 945–981, 2018, doi:https://doi.org/10.1007/s11277-017-4903-y.
- [15] G.J.S. C. Rosewarne, B. Bross, M. Naccari, K. Sharman, "High efficiency video coding (HEVC) test model 16 (hm 16) improved encoder description update 9," in document: Jctvc-ab1002, joint collaborative team on video coding (jctvc) of itu-t sg16 wp3 and iso/iec jtc1/sc29/wg11 28th meeting, 15–21, 2017.
- [16] M. Wien, "High Efficiency Video Coding: Coding Tools and Specification," Springer-Verlag, 2015.
- [17] H.Y.W. C.H. Mai, Y.C. Huang, "Cross-layer adaptive H.264/AVC streaming over IEEE 802.11e experimental testbed," in *IEEE 71st Vehicular Technology Conference*, 1–5, 2010.
- [18] D.W. Q. Chen, "Delay-rate-distortion model for real-time video communication," *IEEE Trans. Circuits Syst. Video Technol.*, **22**(12), 1376–1394, 2015, doi:https://doi.org/10.1109/TCSVT.2015.2389391.
- [19] Y.I. G. Kokkonis, K.E. Psannis, M. Roumeliotis, "Efficient algorithm for transferring a real-time HEVC stream with haptic data through the internet," *J. Real-Time Image Process.*, 343–355, 2016, doi:https://doi.org/10.1007/s11554-015-0505-7.
- [20] X.S. C. Han, M. Dianati, R. Tafazolli, R. Kernchen, "Analytical study of the IEEE 802.11p MAC sublayer in vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, **13**(2), 873–886, doi:https://doi.org/10.1109/TITS.2012.2183366.
- [21] V. Srivastava, M. Motani, "Cross-layer design: a survey and the road ahead," *IEEE Commun. Mag.* **43**(12), 112–119, 2005, doi:https://doi.org/10.1109/MCOM.2005.1561928.
- [22] G.J. van R. M.J. Booyesen, S. Zeadally, "Survey of media access control protocols for vehicular ad hoc networks," *IET Commun.*, **5**(11), 1619–1631, 2011, doi:https://doi.org/10.1049/iet-com.2011.0085.
- [23] *Intelligent transport systems (ITS); access layer specification for intelligent transport systems operating in the 5 GHz frequency band*, ETSI EN 302 663, 1.2.1, 1–24, 2013.
- [24] A. Festag, "Standards for vehicular communication—from IEEE 802.11p to 5G," *E&I, Elektrotech. Inf.Tech.* **132**(7), 409–416, 2015.
- [25] *IEEE standard for wireless access in vehicular environments (WAVE)—multi-channel operation*, *IEEE Std.* **1609.4**, 1–89, 2010, doi:https://doi.org/10.1109/IEEESTD.2011.5712769.
- [26] R. Zhang, L. Cai, J. Pan, "Resource Management for Multimedia Services in High Data Rate Wireless Networks," Springer-Verlag, 2017.
- [27] *IEEE standard for wireless access in vehicular environments (WAVE)—multichannel operation*, *IEEE Std.* **1609.4–201**, 1–89, 2011, doi:https://doi.org/10.1109/IEEESTD.2011.5712769.

Taxonomy of Security Techniques for Routing Protocols in Mobile Ad-hoc Networks

Kartit Zaid^{*1,2}, Diouri Ouafaa²

¹Mohammed V University in Rabat, Presidency of Mohammed V University, Innovation Center, Rabat, 10102, Morocco

²Mohammed V University in Rabat, Computer Science Department, Mohammadia School of Engineering, Rabat, 10102, Morocco

ARTICLE INFO

Article history:

Received: 25 November, 2021

Accepted: 26 January, 2022

Online: 11 March, 2022

Keywords:

MANETs

Security Routing

Authenticity

Trust

ABSTRACT

The Nodes equipped with wireless technology cooperate in an autonomous and instantaneous way to form a mobile ad hoc network. It turns out that several factors make this type of network vulnerable to various security threats. Considering the sensitivity of user data routed through nodes, routing security should be a priority in mobile ad hoc networks (MANET). Techniques and schemes have been proposed to secure the basic routing protocols in order to guarantee the availability of information routing services between network nodes. The majority of the solutions presented in the literature belong to two categories, namely those that use cryptographic techniques and those that use trust schemes. Given the characteristics of MANET networks, we need approaches that guarantee a level of honesty of the nodes to prevent possible routing attacks from malicious nodes. This study presents the security extensions of the basic routing protocols AODV, DSR and DSDV. A first part is devoted to extensions based on cryptography and a second part introduces extensions using trusted systems. Then we discussed and analyzed them while drawing up a comparative table to measure the effectiveness of the mechanisms used as well as the limits and strengths of each proposed extensions. In this study, we conclude that a new trust model that combines an access strategy with lightweight techniques must be developed to ensure honest node behavior can be a key to securing the routing protocol in MANET.

1. Introduction

Mobile Ad hoc Network is becoming an interesting research field as it offers great flexibility and a fast and fluent dynamic implementation. Indeed, mobile ad hoc network MANET (Fig. 1) is an autonomous system consisting of a collection of nodes that are interested in communicating via a wireless link. In MANET, the node is self-configured without the need for any central administration and can communicate directly with neighbors. But to communicate with out-of-range nodes, it requires the cooperation of other intermediate nodes which act as routers to establish reliable and optimal route [1].

The absence of a communication administration makes the deployment of mobile ad hoc networks easier. But the reliability of routing information exchanged between nodes when establishing routes and maintaining them presents a challenge because these networks have characteristics that make them more vulnerable to attack from malicious node. This last node cannot respect the routing protocol rules by disturbing the routing process by inserting false information in the routing messages, modifying

their content's or simply not cooperating by deleting them. Indeed, several studies in the literature have shown that the nodes are exposed to several threats security of routing protocol and in [2] they has established a taxonomy of attacks detected in the MANET. Although cryptographic techniques have been widely used in routing to protect routing information, such an approach may not be practical for real MANETs due to heavy computational loads and the lack of ability to detect attacking nodes [3].

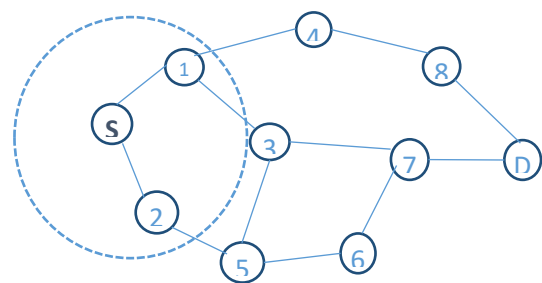


Figure 1: A Mobile Ad hoc Network

*Corresponding Author: Kartit Zaid, Email: zaid.kartit@um5.ac.ma

Nodes continually join and leave networks. It turns out that taking into consideration the notion of "trust" in ad hoc networks is very useful in a highly dynamic environment where nodes must depend on each other to accomplish their common goals. A trust system allows each node of the network to observe and predict the behavior of its neighbors in an efficient manner, with the objective of selecting honestly cooperating nodes.

Therefore, trust-based routing has been seen as an effective measure to manage security threats caused by malicious nodes through detection and isolation of untrusted nodes in the network [3]. Only, to create a reliable communication environment, a distributed trust system that supports network access control and honest cooperation, according to specific routing protocols is needed to help nodes achieve their mission in the mobile network ad hoc.

This paper is organized as follows, section 2 introduces security routing in MANETs, section 3 describe and discuss the different techniques deployed by researchers to solve partially the problem of routing security, then section 4 present the security analyses of extensions studied. Then we conclude our study and discussion in Section 5.

2. Routing protocol and security in MANET

For an ad hoc mobile network, a routing protocol is considered the first part to ensure self-training of such a network. In the literature the researchers did not take into consideration the security aspect that made MANETs networks very vulnerable to attacks. A necessary first action that such a protocol must ensure is to ensure that each node cooperates in an honest way.

Any attack on the routing process disrupts communications between network nodes and can go up to and disrupt the entire network. A second action can be summarized in that it takes into consideration the constraints posed by the scarcity of resources available to nodes such as residual energy, storage capacity.

To provide security services the previous studies in [1][4] have outlined several solutions classified in two categories namely prevention mechanism or detection and reaction mechanism [5][1]. The first one involves encryption techniques particularly to ensure the authenticity and integrity of messages. The latter is used for detecting malicious behavior of dishonest nodes.

Therefore, we can say that a Routing Protocol in MANETs networks can be considered as a communication model whose components and the role of each of them must be specified.

3. Techniques for secure routing in MANET

This paper is an extension of previously published conference paper originally presented in Networking, Information Systems & Security conference NISS19 [1]. This section presented some relevant security extensions of the AODV, DSDV and DSR. We have classified them into two categories, the first approach based on cryptography systems and the second based on trusted schemes.

3.1. Cryptographic techniques

Cryptography systems are widely used in the security of communication networks. It is in the continuity of use of such systems that researchers have applied them to secure routing

protocols in ad hoc networks. In the following we present various cryptography solutions used to secure routing protocol in MANETs. In our previous work published at ACM [1], we has compared and analyzed some extensions of DSR, DSDV and AODV protocols based on cryptographic techniques. In addition to these extensions, we present others in this paper that we consider relevant to solve the problem of routing security in MANET networks. Indeed a common point of these models is that they focus on the importance of identity and decentralized aspect that must be taken into consideration.

Concerning the DSR protocol we cite two extensions which used two different techniques. In the first extension [6] the authors proposed a SecDSR security extension of DSR protocol based on Ariadne protocol which relies on symmetric cryptography that is able to authenticate the source who initiated route discovery process. In fact this extension integrated the message authentication code (MAC) mechanism to provide a point-to-point authentication of routing messages between two nodes in mobile ad hoc network. In the second extension [7] the authors integrated in DSR protocol the sequential aggregate ID-based signatures (IBSAS). In fact, during the route discovery phase to validate the routing information, each intermediate node participates in the calculation of the IBSAS signature conveyed in SRREQ. This allows the destination node to verify the signature in the SRREQ using the IDs of intermediate nodes received by IBSAS verification. ISDSR reduce memory size and compute overhead. Therefore ISDSR helps eliminate dishonest nodes in a route. ISDSR can ensure the validity of route information when a route is constructed.

As for the DSDV protocol, in order to ensure the authentication and the integrity of the messages, extensions have incorporated mechanisms based on asymmetric cryptography. In [8] the authors presented dAN DSDV an extension of the DSDV protocol using a message authentication code based on an asymmetric cryptographic approach to establish an optimal secure path. The certification and validation of the RREQ message are performed at each node when moving from one node to another using a paired shared secret key mechanism in the routing path. The proposed method ensures both the authentication and the integrity of the message with packet loss and delay minimized, throughput and packet delivery rate maximized. To allow mutual authentication between two network nodes in [9] the authors proposed a mutual Hash-MAC-DSDV scheme by making a modification of the DSDV protocol. This scheme adopts a structure of a cluster network where the CH nodes facilitate the registration of nodes in a local channel at first and broadcast it via nodes called base stations to register it in a second called public channel. This process allows the network node to update their routing table to be able authenticate each node requesting communication. Indeed, before transmission, a unidirectional authentication process is performed to verify the legitimacy of each node by matching Hash MAC with their table of MAC addresses. This scheme has demonstrated its effectiveness in terms of attack, detection metrics, power consumption and communication cost.

Attack prevention is considered an effective way to provide a level of security for the AODV protocol. Node identification and key management have been used to strengthen its security. Indeed, in [10] the authors integrated the HiMAC mechanism to secure the

distribution of data in AODV. Based on trust and message authentication the HiMAC's technique prevent an intruder from capturing routing information, which makes it easier for him to carry out attacks by generating a diversion of the route thus causing a black hole attack. This prevents intruders from altering or modifying the number of hops by using signature and encryption of messages at each intermediate node. This approach is based on identity-based cryptography and message authentication code (MAC). In fact, each intermediate node will add its MAC and its timestamp, in addition to its ID to the list of message identifiers. In this scheme, each node shares its public key with other "trusted" nodes only. HiMAC require an efficient public key management system or the key size must be optimal to reduce encryption and decryption times and the size of the data messages is not very large. In [11] the authors proposed a new SAODV protocol scheme by introducing a short digital signature scheme for authentication. The use of the short digital signature aims to gain the same level of security but with more efficiency and less computation. In this scheme another signature assignment mechanism is to use a trusted third party to share secret digital signatures with each node in the network and establish secure communications between two nodes by signature-based authentication. In this scheme the hash function was used to authenticate the hop count.

3.2. Technical Trust

The designers of routing protocols for ad-hoc networks have considered that the intermediate nodes cooperate in the respect of the specificities of said protocol by expressing an honest behavior. This supposed trust allows malicious nodes to easily generate attacks on the routing process. Indeed, malicious nodes can deliberately behave in such a way as to disrupt the content of the packets and consequently disrupt the routing process. Several security mechanisms have been proposed to protect routing information against attacks by malicious nodes, to establish a reliable and efficient communication path. Many approaches and proposals have been proposed to address various trusted ad hoc secure routing schemes.

A malicious node is considered a primary source of attack for basic protocols such as AODV. To eliminate this type of attack in [12], [1] the authors presented a trust mechanism named TDS-AODV. This mechanism has been modified AODV routing protocol to implement the trust model of TSDRP [13] to prevent malicious actions like Blackhole and DoS attacks by calculating the trust value for their neighbor. In this AODV extension, a node makes a routing decision based on the trust values of its neighbor nodes. Finally, two routes are built: the main route with the highest route trust value in the candidate routes and the backup route. This mechanism has proven its ability to eliminate malicious nodes during the construction of the route. This extended protocol above highlights efforts to introduce and improve reliability in mobile ad hoc networks using trusted systems.

To ensure the existence of links to route messages in the MANET in [14] the authors have proposed a Trusted Recover AODV (TR-AODV) model based on the identity authentication scheme and the existing TAODV trust model. Their model requires a node to join the network to be authenticated in two cases where it is the first time or if it has a trust level less than the threshold. This model encourages nodes already registered by an

identity to join the network by resetting their trust level to the threshold value. This will guarantee the availability of routing service and push the nodes to cooperate and behave normally according to the system of trust. If his cooperation is verified, he improves his trust value. Otherwise, it will go to the blacklist. This technique has the impact of good management of malicious nodes by giving them a second chance to participate effectively in the network. Consequently, packet loss is reduced and transmission delay is improved.

In another scheme the addition of data structure seems necessary to route the messages between nodes in order to distribute trust in the network. Indeed in [15] the authors proposed ESTA protocol based on AODV by introducing two new data structures LINK_TABLE and LINK_INFO. Each node serves as a LINK_TABLE mechanism to first record the information contained in the RREQ requests received and in a second to generate a LINK_INFO control message once it is part of a path. The latter makes it possible to update its availability and that of the next node in the LINK_TABLE table of neighboring nodes. This protocol works without a certification authority; however, it combines a system of trust with asymmetric cryptography to ensure the integrity and authentication of control messages. To satisfy the authenticity of the messages, it encrypts the node identity and the timestamp value by the private key of the source node. Another data structure called SOURCEINFO is used to store the level of the trust value vis-à-vis the network nodes. This value is communicated by another control message SEND DELIVERY_INFO. All of this impacts communication delay due to the overhead introduced by these control packets and data structures. Another disadvantage is that the trust system is limited to direct observations. Indeed, in the absence of interaction between two nodes or sharing of trust with other network nodes, a malicious node being out of range can respond to its RREQ message leaving the source node unable to make a good decision. It is desirable to consider indirect observations, especially since they have added control messages and data structures that can help ensure good management of the trust.

The selection of the path among others is always based on the number of hops to reach the destination. But in [16] the authors have proposed the ReTE-AODV based protocol which uses the level value of trust as a path selection parameter to guarantee the honest commitment of the nodes which will transmit the data packets. The proposed algorithm routes the messages not by the shortest route, but by selecting a reliable trustworthy route that consumes little power and has a level of trust for sending the packets. To do this the trust value is obtained from direct and indirect trust. The calculation of the trust value is based on the direct and indirect trust and to refine this value they adopted the Bayesian method. The same thought to change the paradigm of the choice of the route in [17] the authors have proposed a multi-factor routing strategy named EOSR protocol. This extension of the AODV protocol adopts a distributed trust model that helps to detect and isolate malicious attacks. In their solution the choice of the route is based on three metrics that is residual energy, trust level and number of hop count. The behavior of the node is used to calculate the level trust. In order to improve the accuracy of the trust value they required to obtain the indirect trust from the common adjacent nodes. However, to filter out false notes from malicious nodes, any indirect trust collected by an adjacent node

must have excluded false notes that exceed the deviation threshold. Trust and energy information is added in routing messages without increasing communication traffic. As for path selection, they defined a new metric called total path cost taking into account the confidence value, the residual energy and the number of hops count. One of the major advantages of this extension is that it takes into account the dynamic change of the network topology.

In order to demonstrate a level of trust, several measures must be taken into consideration. This is why in [18] the authors proposed a system called SCOTRES, integrated into DSR, based on trust to secure the routing of the mobile ad hoc network. This system advances the intelligence of network node by applying five new measures. The energy metric to measure the level of cooperation expected from each node. The dynamic topology metric anticipate the change position of node. The channel quality metric gives an idea of the reliability of the node. The reputation metric assesses the cooperation of each participant in a specific network operation, detecting attacks, while the trust measure estimates overall compliance, protection against combinatorial attacks. In another view in [19] the authors have proposed the DSR Enhanced Trust (TEDSR) based on trust to establish stable and reliable routes. It includes the payment systems in a trust-based routing protocol. The goal is to establish the stable route to reduce the liability of route breaks. The enhanced DSR protocol establishes the best path that can meet the requirements of the source node including energy, trust level and route length, and incorporated this information's into routing messages (RREQ, RREP).

Finally in [20] the authors present a model based on the Blockchain structure (BATM) to ensure authentication and trust in sensor networks. The Blockchain is used to facilitate the management of public keys, digital signature and peer information. Consequently, each node of the network has the possibility of validating the information on all the other nodes of the network. The BATM module includes a trust model called Human-based Knowledge-Based Trust (HKT), which is based on human behavior to maintain a reputation level for each node. It uses the payloads contained in the Blockchain to measure the behavior of each node. In this way, it ensures that a node cannot deceive others by falsifying data or impersonating someone else. Thus, it assesses trust, without the need for a trust center.

4. Security analysis, discussion, and comparative study

The reliability of the information diffused through the control messages in the different phases of routing is the basis for a secure routing protocol solution. In fact, an ad hoc network is a network that necessitates a degree of cooperative trustworthiness where each node must respect the specifications of the routing protocol.

For the first category based on cryptography, the various solutions deployed above are dedicated to the authenticity of the nodes and the integrity of the routing messages. Given the characteristics of the MANETs networks, the proposed solutions go toward techniques that combine symmetric cryptography systems and hash functions. On the other hand, asymmetric cryptography systems secure efficiently but are not easily adaptable for this type of network. Several mechanisms are used to prevent active attacks that affect the update of messages and

their authenticity as digital signature, message authentication code (MAC), one-way hash functions, hashed MAC (HMAC), or a combination of these techniques [1]. To secure validity of the route information when a route is constructed several solutions have been developed such as:

In dAN EFFICIENT DSDV the destinations can verify the integrity of the message using a hash chain calculated at the time of broadcasting of the RREQ. As for the authentication it is ensured by including the identity of each participating node in the hash code. Therefore, this technique provides both authenticity and integrity of routing packets. We consider this mechanism efficient since it proves the honesty of the intermediate nodes to establish the secure and guard the path establishing. In SecDSR, the authors used the TESLA hash chain to calculate the code MAC as well as a hash accumulation including the identity of each intermediate node to prevent corruption of path information's. ISDSR is based on digital signature to trust the information communicated by a node. In effect, indeed to participate in the creation of the route it is necessary to have an identifier and a secret key and this to prove a first level of trust towards others. Only the drawback of this approach is the fact that it requires the existence of a server for managing identifiers and keys for each node. In addition, it requires each node to reserve a memory space to store the keys. The Hach-MAC-DSDV scheme adopts a network zoning to facilitate the management of the network nodes. Each zone is administered by a CH zone manager and a local chain which records the hash of the MAC address of the nodes registered in the network. A public chain is established between the different CH. this allows the creation of a decentralized pseudo-administration to ensure the authentication of the participating nodes by creating a secure authentication key for each node. A major vulnerability of this model is that it allows each node to register with its MAC address, which is not enough to prevent a malicious node from subsequently carrying out an attack once it has access to the network. HiMAC is one of the first techniques which is based in the first place on cryptographic tools based on the identity of the nodes and the MAC code and in a second on the confidence of the participating nodes in the dissemination of messages in the network. HiMAC can be efficient with a light PKI and with a powerful identification system.

Although cryptographic techniques have been widely used to protect routing information based on checking the authenticity and integrity, such an approach is not sufficient to secure routing in MANET networks. We observe in the whole of the preceding diagram that the identity of the nodes plays a primordial role in ensuring the security of the routing protocols in an ad hoc network. With this in mind, other extensions belonging to a second category based on trust have emerged by introducing the notion of trust to help nodes to observe and predict the behavior of neighboring nodes efficiently [3]. Trust-based systems extend the level of node cooperation that goes beyond simple verification. Indeed, in the ad hoc network, each node measures a trust level of the nodes before entering into interaction with them. Each extension is based on a scheme to assess the confidence of the nodes. Direct observations, recommendations from others, and other factors can change a node's trust level [21]. Most of the extensions studied in this paper consider that path selection should take into consideration the trust level of and not just the number of hops. Such a trust level is not limited to the behavior of the nodes, but it's extended to other

factors such as residual energy, the quality of the available channel, and the dynamic aspect of the topology [22].

The TR-AODV protocol considers that the availability of a cooperating node is necessary, so these authors have adopted the restoration of the reputation of nodes via a node authentication mechanism before accessing the network. We believe that this action is very necessary to eliminate dishonest nodes from sharing the network and to get others to cooperate honestly. The EOSR protocol is a multi-factor routing protocol based on a distributed trust model that takes into account the dynamic aspect of the network. As for the SCOTRES protocol, it integrates the intelligence of the nodes of the network by applying five new measures to assess the level of confidence of the nodes. It exhibits the best power and load balancing behavior, provides the highest level of security, and handles certain types of attacks that other systems cannot counter. Another very important measure is introduced in the ESTA protocol which does not require a certification authority. BATM provides an easy way to manage trust in decentralized Blockchain-based networks and takes more parameters into account in assessing trust and reputation.

In summary, in the first part of our table below we notice that the cryptography-based extensions are unable to preventing all security threats. We also concluded [1] that these protocols should not consider all network nodes as trusted. Therefore, the way they build the roads is to be reviewed. Another point that makes routing security more complex is the dynamically changing network topology, which makes the paths breakable and unstable. In addition, we can say that security must be ensured for the entire system because a single weak point can give the attacker the ability to access the system and perform malicious tasks. In second part of our table we congn that each extensions based trusted system provides a technique to reduce the drawbacks of cryptography. And others that allow routing protocols to be adapted to the characteristics of Ad Hoc networks, such as the notion of autonomous and distributed management, the measurement of the quality of interaction between network nodes.

Indeed, we have identified three essential points to take into consideration in our future work to secure a routing protocol in MANET's networks. At first, we will develop a new distributed

trust system in which trust will not only depend on authentication, we will define a metric to measure the behavior of the nodes, each node must show its honesty in the forwarding process of messages. Then, an efficient method to calculate the index of the reputation of the nodes seems to us necessary. We will use this index to define the level of trust. As a second step, we also propose a revision of the RREQ broadcast approach in which we will limit access to routing information to nodes with a certain level of trust. In fact, the new nodes will have a low level of trust. By cooperating with honest behavior, their level of trust becomes great. We recall that in ad hoc networks the nodes are considered trustworthy when they arrive. In our approach, we will reverse this principle by applying the method described above [1].

5. Conclusion

This paper reviewed a taxonomy extension security grouped by technics used in various researches to secure the routing protocols. Different approaches based on cryptography and / or on trusted systems are proposed to prevent routing attacks in MANET networks. These approaches try to provide an optimal path composed of secure node by implementing different mechanisms in existing routing protocols. We have seen the limit of cryptography-based solutions either in terms of adapting to the resource constraints that this type of network suffers or in responding to security objectives. As for trust-based protocols, most extensions consider that the path selection should take into consideration the level of trust and not just the number of hops in the path selection. Such a trust level is not limited to the behavior of the nodes but also to other factors such as residual energy, the [1] quality of the available channel and the dynamic aspect of the topology. But we have observed other limitations which are more precisely linked to the identification system and the characteristics of Ad hoc networks. Finally, this study will help us in our future work to introduce other technologies based on the hardware properties of nodes to facilitate the management of the identities of the nodes that we consider the security gate of these protocols.

Conflict of Interest

The authors declare no conflict of interest.

Table 1: Comparison of secure extensions routing protocols in Mobile Ad Hoc network

Extension	Attacks	Parameters and Mechanisms	Advantages	Disadvantages
dAN-DSDV	Malicious	HMAC, ID	Packet drop excluded Delay time decreased with increasing the security level	energy consumption overhead
SecDSR	Malicious	Tesla, hash chain, ID	More security than DSR	increases routing overheads congestion in the network
ISDSR	Malicious	IBSAS, aggregate signatures	decrease the memory size and the computational overhead better latency	management server required
Hash-MAC-DSDV	Sybil, Dos Eavesdropping	lightweight Hash-MAC clustering local and public chain ,one-way-hash authentication	Legitimacy of node guaranteed Better attack detection rate minimal resource consumption	difficult to implement

HiMAC-AODV	Tampering data, replay attack	MAC, Digital Signature , trust schemes	higher success ratio, less hop count, smaller packet queue size secure routing dynamically	high overhead high processing overhead
SAODV	Blackhole Grey hole	Short digital signature, Hash function	reducing the computational power	signature distribution center required
TR-AODV	Replay Selfish	Authentication and Reputation Restoring and Increasing Chances for Normal Communication (Availability)	Data loss rate reduced , Incentive for node cooperation Better network Improved network access ,Delay reduced Communication quality	Authentication and reputation restoration scheme load to consider
ESTA	Blackhole	Trust level, Identity encryption control packages add	not use CA or additional processing at intermediate	Limited to direct observation, Consumes more resources additional delay
ReTE-AODV	Malicious	Trust level energy consumption	Good level of security with a better packet delivery ratio and reduced average end-to-end latency.	Consumes more energy, routing packet overload
EOSR	Malicious	Residual energy, CCP full path cost , Distributed trust model	Best performance in terms of delivery rate, throughput with average energy consumption. Support the dynamic aspect	End-to-end delay is not included in the performance metric
TEDSR	Blackhole	three levels of trust Payment report	Path securing by validation of intermediate nodes by the source node identify fraudulent nodes accurately and quickly without false accusations or missed detections	Transfer-based trust level calculation method is not sufficient to measure the behavior of nodes. Not tested in the presence of malicious nodes.
SCOTRES	flooding Blackhole, link-spoofing	Uses five metrics: residual energy, channel quality, topology, reputation and trust levels	Best behavior in terms of energy and load balancing, provides the highest level of security and handles certain types of attacks that other systems cannot counter.	Consumes more resources Additional traffic overload
BATM	DOS	Blockchain PKI Trust level	Simple way to manage trust in decentralized Blockchain-based networks. takes more parameters into account when assessing trust and reputation levels	First block creation Lack of performance measurement Storage space required for the Blockchain

References

[1] Z. Kartit, O. Diouri, "Security extension for routing protocols in Ad hoc mobile networks: A comparative study," ACM International Conference Proceeding Series, Part F1481, 2019, doi:10.1145/3320326.3320403.

[2] N.A. Noureldien, "A novel taxonomy of MANET attacks," Proceedings of 2015 International Conference on Electrical and Information Technologies, ICEIT 2015, 109–113, 2015, doi:10.1109/EITech.2015.7162947.

[3] M.S. Pathan, N. Zhu, J. He, Z.A. Zardari, M.Q. Memon, M.I. Hussain, "An efficient trust-based scheme for secure and quality of service routing in MANETs," Future Internet, 10(2), 2018, doi:10.3390/fi10020016.

[4] A.K. Abdelaziz, M. Nafaa, G. Salim, "Survey of routing attacks and countermeasures in mobile ad hoc networks," Proceedings - UKSim 15th International Conference on Computer Modelling and Simulation, UKSim 2013, 693–698, 2013, doi:10.1109/UKSim.2013.48.

[5] K. Vijayakumar, K. Somasundaram, "Study on reliable and secure routing protocols on Manet," Indian Journal of Science and Technology, 9(14), 2016, doi:10.17485/ijst/2016/v9i14/84433.

[6] M.K. Hameed, F.J.Abd-Razak, "A Secure dynamic source routing protocol for mobile ad hoc networks," journal of kerbala university 15, issue 4, 32-41-2017

[7] K. Muranaka, N. Yanai, S. Okamura, T. Fujiwra, "ISDSR: Secure DSR with ID-based sequential aggregate signature," ICETE 2016 - Proceedings of the 13th International Joint Conference on e-Business and Telecommunications, 4(Icete), 376–387, 2016, doi:10.5220/0006001003760387.

[8] Ch. Anusha, E. Laxmi Lydia, T. Pavani, Ch. Usha Kumari, M. Ilayaraja Kanagaraj Narayanasamy, "dAN efficient dsdv routing in mobile networks through symmetric cryptographic method," , Journal of critical reviews ISSN- 2394-5125 VOL 7, ISSUE 10, 2020. doi:10.31838/jcr.10.31.

[9] M. Adil, M.A. Jan, S. Mastorakis, H. Song, M.M. Jadoon, S. Abbas, A. Farouk, "Hash-MAC-DSDV: Mutual Authentication for Intelligent IoT-Based Cyber-Physical Systems," IEEE Internet of Things Journal, 4662(c), 1–11, 2021, doi:10.1109/JIOT.2021.3083731.

[10] K. Mershad, A. Hamie, M. Hamze, "HiMAC: Hierarchical Message Authentication Code for Secure Data Dissemination in Mobile Ad Hoc Networks," International Journal of Communications, Network and System Sciences, 10(12), 299–326, 2017, doi:10.4236/ijcns.2017.1012018.

[11] M.T. Abbas, M.A. Khan, A. Khaliq, N.A. Saqib, J. Ahmad, S. Rehman, "Secure AODV Protocol for Mobile Networks Using Short Digital Signatures," Proceedings - 2017 International Conference on Computational Science and Computational Intelligence, CSCI 2017, 645–650, 2018, doi:10.1109/CSCI.2017.111.

[12] R. Feng, S. Che, X. Wang, N. Yu, "A credible routing based on a novel trust mechanism in Ad Hoc networks," International Journal of Distributed Sensor Networks, 2013, 2013, doi:10.1155/2013/652051.

[13] A. Aggarwal, S. Gandhi, N. Chaubey, K.A. Jani, "Trust based secure on demand routing protocol (TSDRP) for MANETs," International Conference on Advanced Computing and Communication Technologies, ACCT, 432–

438, 2014, doi:10.1109/ACCT.2014.95.

- [14] J. Liu, S. Huan, "Trust recovery model of Ad Hoc network based on identity authentication scheme," AIP Conference Proceedings, 1839(May), 2017, doi:10.1063/1.4982566.
- [15] D. Singh, A. Singh, "Enhanced secure trusted AODV (ESTA) protocol to mitigate blackhole attack in mobile Ad hoc networks," Future Internet, 7(3), 342–362, 2015, doi:10.3390/fi7030342.
- [16] Priya Sethuraman, N. Kannan, "Refined trust energy-ad hoc on demand distance vector (ReTE-AODV) routing algorithm for secured routing in MANET," Wireless Networks, 23(7), 2227–2237, 2017, doi:10.1007/s11276-016-1284-1.
- [17] T. Yang, X. Xiangyang, L. Peng, L. Tonghui, P. Leina, "A secure routing of wireless sensor networks based on trust evaluation model," Procedia Computer Science, 131, 1156–1163, 2018, doi:10.1016/j.procs.2018.04.289.
- [18] G. Hatzivasilis, I. Papaefstathiou, C. Maniavas, "SCOTRES: Secure Routing for IoT and CPS," IEEE Internet of Things Journal, 4(6), 2129–2141, 2017, doi:10.1109/JIOT.2017.2752801.
- [19] R. Pricilla, T.R. Vadhavathy, "TRUSTED ENHANCED DSR FOR IMPROVING PAYMENT SCHEME IN MWN," (December), 2–6, 2014.
- [20] A. Moinet, B. Darties, J.-L. Baril, "Blockchain based trust & authentication for decentralized sensor networks," (June), 2017.
- [21] Z. Hao, Y. Li, "An adaptive load-aware routing algorithm for multi-interface wireless mesh networks," Wireless Networks, 21(2), 557–564, 2015, doi:10.1007/s11276-014-0804-0.
- [22] K. Kundu, C. Chowdhury, S. Neogy, S. Chattopadhyay, "Trust aware directed diffusion scheme for wireless sensor networks," Proceedings - 4th International Conference on Emerging Applications of Information Technology, EAIT 2014, 385–391, 2014, doi:10.1109/EAIT.2014.68.

Power Management and Control of a Grid-Connected PV/Battery Hybrid Renewable Energy System

Othmani Mohammed*, Lamchich My Tahar, Lachguar Nora

Intelligent Energy Management and Information Systems Laboratory, Physics Department, Cadi Ayyad University, Faculty of Sciences Semlalia of Marrakech, Marrakech, 40000, Morocco

ARTICLE INFO

Article history:

Received: 25 November, 2021

Accepted: 26 February, 2022

Online: 11 March, 2022

Keywords:

Energy Management Strategies

Grid-Connected

Hybrid Renewable Energy System

Peak Shaving

Power Flows Control

ABSTRACT

This paper presents novel Energy Management Strategies (EMSs), and the control of a Grid-Connected Hybrid Renewable Energy System (GCHRES). The GCHRES describes a Photovoltaic Generator (PVG) and a Battery-Based Storage System. Both are tied to the Common Coupling Point (CCP) through a reversible three-phase inverter. The CCP combines the Utility Grid (UG) and an AC house load supplied primarily by the PVG. The UG supports the PVG in case of deficit. The battery is designed for peak shaving application to avoid the UG subscription power exceeding. Due to the weakness of the battery-bank, all EMSs aim to ensure continuity of supply while preserving the battery-bank from overcharges and high depth of discharges. This being said, the reducing of its lifespan would be avoided. Furthermore, These EMSs aim also to reduce the monthly UG customer energy bill. They differ depending on the type of metering with the UG. The energy balance within the system is ensured by controlling the battery and the PVG powers, as well as the power exchange with the between the inverter and the UG. The performances of the GCHRES, under different metering cases, and under several operation modes, were verified in MATLAB/Simulink based on real solar irradiation and temperature profiles data corresponding to the region of Marrakech, Morocco. Results in terms of demand meeting, DC-Bus voltage regulation, global system stability and power references tracking are presented in this paper.

1. Introduction

Nowadays, the energy demand is permanently increasing, causing an energy crisis due to the depletion of fossil energy sources such as coal, oil and natural gas. Therefore, the price of electricity from centralized generation keeps increasing as well. In addition, the environmental deterioration of the planet due to Greenhouse Gases emission, as well as the technological development of energy control, have made the integration of renewable energy sources more attractive for professionals as well as for individuals. These clean and abundantly available renewable power sources such as solar and wind powers, which fall under the scope of Decentralized-Generation, are currently more profitable and widely exploited, and can be used either in Grid-connected or autonomous modes. The application of PV systems has become more common in developed and growing countries, and their performances are better in high solar irradiation areas, and can provide enough power if properly operated. Due to the intermittency of climatic conditions, the

output power of the Photovoltaic Generators (PVGs) is affected. This latter have enormous energy potential, even if their efficiency is relatively low (25% to 30%). Under these variable conditions, the power delivered changes according to the voltage imposed on their terminals. In order to take advantage of the fully PVGs potential, the Maximum Power Point (MPP) of the Voltage-Power curve must be reached for any solar irradiation and temperature value. Maximum Power Point Tracking (MPPT) algorithms have been developed to achieve this purpose. Several works have been carried out for MPPT algorithms development, among which stand out [1]-[4], and which differ according to a compromise between complexity and desired performances. An MPPT using Adaptive Fuzzy Logic control for Grid-connected Photovoltaic systems is presented in [5]. These MPPT algorithms are divided into two families: on the one hand those based on power derivation method such as Incremental Conductance (INC). On the other hand those based on voltage/current feedback such as the Perturb and Observe (P&O) [6]. A comparison of the dynamic and the speed of different MPPT algorithms is presented

*Corresponding Author: Othmani Mohammed,
othmani.mohamed.mmed@gmail.com

www.astesj.com

<https://dx.doi.org/10.25046/aj070204>

in [7]. The most commonly used MPPT algorithm is P&O due to its simple implementation and good performances [8], [9].

One of the major advantages of PVGs is their smooth integration via existing power converters. Unfortunately, their output power is highly dependent on weather and climatic conditions. Therefore, perfect services to UG and loads, requiring constant power profiles, cannot be guaranteed. Consequently, PVG alone will not lead to satisfying results. Hence, it is either Grid-connected or associated with a storage system (Hybridation), or even with both simultaneously, to ensure supplying continuity. The storage device can be a battery-bank, fuel-cell/electrolyzer system, flywheel, super-capacitor, compressed air... and the way the energy is stored depends on the application.

As said before, solar systems can be either autonomous, or connected to the Utility Grid (UG) which represents the purpose of this paper. Grid-connected solar systems can operate either for total, or surplus sale in case of self-consumption. In both cases, a certain power will be destined to injection into the UG. Total sale system is beneficial for individuals in countries where the selling price of solar energy is higher than UG supply energy price. Other countries grant self-consumption bonuses, thus it is more beneficial to maximize the consumption of the produced solar energy, and to sell the surplus to the UG after that. Technically, the difference between these two systems lies on the metering solution. Figure 1 shows two metering solutions that can be adopted for total sale and for surplus sale. In case of total sale as shown in Figure 1(a), a no-consumption meter can be added between the DC/AC converter and the Common Coupling Point (CCP), to avoid charging the battery with the low price energy of the UG, before reselling it to the latter with higher injection price. Regarding self-consumption case, UG provides power when PVG generation is either zero or unable to satisfy the load demand. However, monthly energy bill of UG subscriber increases with the increase of its subscription power. As a result, peak shaving would be an advantageous application of the storage system for the UG subscriber. In fact, the kWh in many countries is more expensive in peak hours where energy demand is high. Peak shaving will reduce the power requested from the UG in this time period, leading to a reducing of the subscription power and the energy bill as well. It will be also beneficial in the perspective of eliminating penalties due to subscription power exceeding. Peak shaving would be beneficial also for the UG and for environment. At the UG level, the decrease in power demand reduces the risks of grid congestion, reducing as well the necessity of calling the backup centralized stations. For example, in Morocco, the kWh produced is still very polluting, given that 61% of electricity is produced from coal centralized stations [10]. Thus, at the environmental level, less demand from these stations means less Greenhouse-Gases emissions. An operating power and discharge time based comparison of different storage devices is presented in [11], and batteries are most suited for peak shaving applications. However, this storage system requires accurate regulation of charge/discharge currents within manufacturer specified range in order to increase its lifetime. Adding constraints of State Of Charge (SOC), which must be kept in a recommended range, to avoid creating harmful irreversible reactions in the battery electrodes, and hence decreasing its lifespan. Sizing properly the

battery-bank is essential, in the aim of getting operational for peak shaving, as soon as the UG power requested reaches its maximum allowed power known as the subscription power. Peak shaving via batteries in France had been studied in [12], and inspired the battery-bank sizing method presented in this paper. Figure 2 shows some metering cases that can occur for Grid-connected solar system in self-consumption mode and which are adopted in Morocco, and will constitute the basis of the development of the proposed Energy Management Strategies (EMSs) in this paper. Whatever the meter is, if solar production is either zero or insufficient to meet load demand, the metering index increases to give the electrical energy consumed in kWh to the distributor, since the deficit is provided by the UG. When the output power of the PVG exceeds the one requested, the metering index works according to the metering type:

- Digital Metering (DM) Figure 2(a): by injecting into the UG, the consumption metering index start increasing as if the individual has consumed the injected energy (concerns around 1 million Moroccan UG subscribers) [13].
- Irreversible Electromechanical Metering (IEM) Figure 2(b): by injecting into the UG, the metering index does not change because the rotation of the disc in the opposite direction is prohibited. Energy is injected and consumed in the neighborhood for free.
- Reversible Electromechanical Metering (REM) Figure 2(c): during injection, the meter index will subtract the injected energy because the rotation of the disc in the opposite direction is allowed.

It is clear that the last metering system is the most favorable for the UG subscriber.

In this paper, battery integration with the aim to realize peak shaving, throughout the different metering cases presented in Figure 2, will be studied, by managing the power flows within the Grid Connected Hybrid Renewable Energy System (GCHRES). This paper deals with a Grid-connected solar system corresponding to a self-consumption operation, and as part of fair net-metering; 1kWh given for one kWh delivered [13], without taking into account solar energy selling and UG energy buying prices. Hence, three EMSs are proposed, depending on the metering type and on the grid injection limitation modes. Their main objective is to permanently satisfy the house loads demand. The battery charge and discharge will be carried out while respecting its technical constraints, related to its SOC and its maximum charging/discharging powers. It is noticed that the system studied do not contain dump load. The first EMS (EMS 1) targets a system equipped with either reversible or irreversible electromechanical meter, when injection into the UG is allowed without limitation. Therefore, the PVG is operating globally under the control of the MPPT algorithm. The second EMS (EMS 2) targets a system with same metering types, but in grid injection limitation case. The output power of the PVG is limited in certain cases by the mean of a limitation power point tracking algorithm, called LPPT and presented in [14]. Different LPPT control schemes are presented in [15]. The two EMS presented above are combined in one flowchart. The third EMS (EMS 3) is destined to digital metering where the subscriber will suffer financial losses in the event of an injection into the UG. Injection should be avoided at all cost. The output power of the PVG is limited in

certain cases by the mean of the LPPT control. The three EMSs are presented in detail in a following section. Performances of the proposed GCHRES under the supervision of the different EMSs are simulated in MATLAB/SIMULINK using real weather data corresponding to the region of Marrakech in Morocco.

This paper is structured as follow: section II presents a description of the proposed GCHRES and the control of its different components. Section III presents the battery-bank sizing method and the proposed EMSs with their different operating modes. Section IV presents and discusses the results obtained by simulation on MATLAB/SIMULINK. Finally, section V summarizes the results obtained in a conclusion.

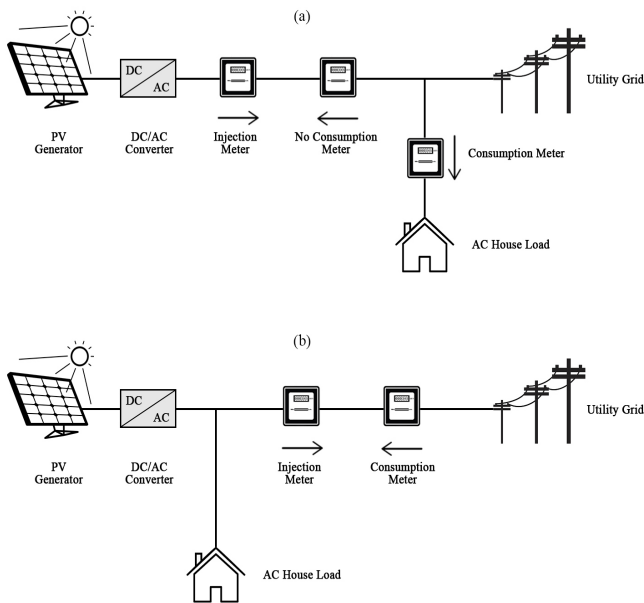


Figure 1: Metering solution for Grid-Connected solar systems (a) total sale (b) surplus sale

2. System Modelling and Control

The proposed system constituting the GCHRES is illustrated in Figure 3. Its architecture based on two buses (DC and AC buses), includes a PVG, considered as the main power source intended to meet the variable demand of a house located in Marrakech, Morocco. The UG is the main backup source in the event of a PVG power deficit. A battery-bank is used for peak shaving application. The PVG and the battery, via their respective power converters, are connected to the DC-Bus. As the DC-Bus voltage is higher than the PVG voltage variation range, the interfacing between these two elements is realized via a DC-DC Boost converter. This solution makes it possible to reduce the number of solar panels constituting the PVG and thus to reduce the initial investment cost. The battery-bank voltage can be kept lower than the DC-Bus Voltage, by realizing the charge and discharge operations through a bidirectional DC/DC Buck-Boost converter (BBDC). The converter operates in Buck mode during the charge, and in Boost mode during the discharge, thus making it possible to limit the number of batteries constituting the battery-bank. The DC-Bus output is connected to the CCP combining the UG and the AC-house via an IGBT based three-phase reversible

inverter called Alternative Side Converter (ASC) in this paper. A state based supervisory controls the power flows within the GCHRES.

It is important to notice that in this paper, expressions including powers related to the battery, UG, PVG and the Load are presented in a way that respects the powers convention sign in MATLAB/SIMULINK. It means that all powers related to the battery charging process are positives and all the ones related to its discharging are negatives. Concerning the UG, all powers related to injection are positives and all the ones related to its consumption are negatives. Unidirectional load and PVG powers are always positives. The detailed description of each part of it, as well as the proposed EMSs will be explained in futures sections

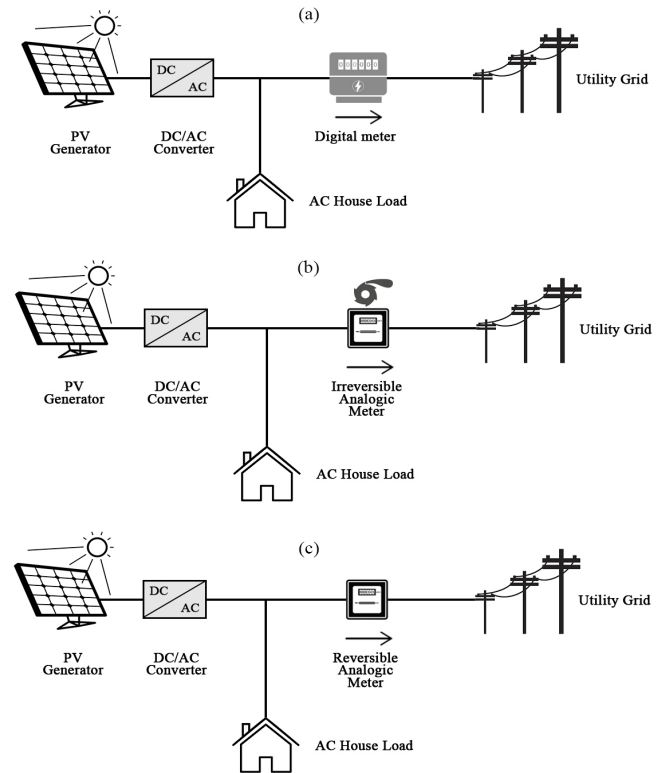


Figure 2: Metering cases for Grid-Connected solar systems in Morocco (a) DM (b) IEM (c) REM

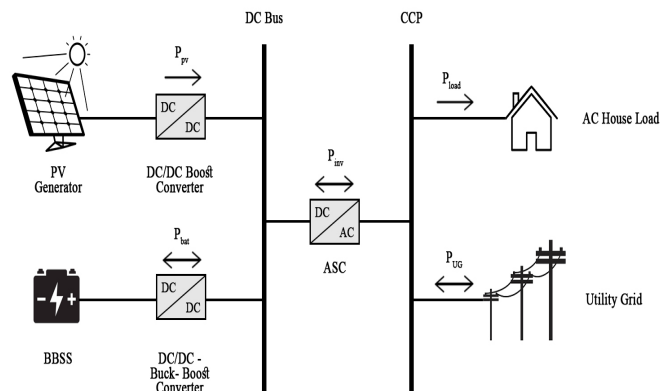


Figure 3: Architecture of the proposed GCHRES

2.1. Photovoltaic Generator Mathematical Modelling

A photovoltaic array is the association of several photovoltaic cells. Many mathematical models of photovoltaic cell have been developed to represent their highly nonlinear behavior, caused by semiconductor junctions constituting the basis of their construction. The single diode representation presented in Figure 4 is adopted in this paper, considered as the most commonly used model [16]. The PV cell output current I_{cell} [A] is given by (1).

where: I_{ph} Photo-Current depend on solar radiation and cell temperature [A]; I_d Diode Current [A]; I_{sh} Shunt Resistance current [A].

The photo-current I_{ph} [A] is given by (2).

Where: I_{sc} Short-Circuit current of the cell under STC [A]; K_T Temperature coefficient of the short-circuit current [A /K]; T_c Temperature of the cell [K]; T_{cref} Reference Temperature of the cell [K]; G Solar Irradiation [W/m^2]; G_{ref} Reference Solar Irradiation [W/m^2].

The current I_d [A] passing through the diode is given by (3), where I_s [A] represents the Diode Reverse Saturation Current given by (4), and I_{rs} [A] given by (5) represents the Diode Reverse Saturation Current under STC.

$$I_{cell} = I_{ph} - I_d - I_{sh} \tag{1}$$

$$I_{ph} = [I_{sc} + K_T (T_c - T_{cref})]G/G_{ref} \tag{2}$$

$$I_d = I_s [e^{(q(V_{cell} + R_s I_{cell})/KT_c A)} - 1] \tag{3}$$

$$I_s = I_{rs} (T_c/T_{ref})^3 e^{\frac{qE_g(T_{ref}^{-1} - T_c^{-1})}{KA}} \tag{4}$$

$$I_{rs} = \frac{I_{sc}}{\frac{qV_{oc}}{eK.N_s.A.T_c} - 1} \tag{5}$$

where: I_s Diode Reverse Saturation Current [A]; q Electron Charge ($1.602 \times 10^{-19}C$); V_{cell} Cell Voltage [V]; R_s Cell Serie Resistance [Ω]; I_{cell} Cell Output Current [A]; K Boltzmann Constant ($1.38 \times 10^{-23}J/K$); A Cell ideality factor dependent on PV technology; E_g Gap energy of the semiconductor used in the cell ($1.1eV$ for Silicon) [eV]; V_{oc} Cell Open Circuit Voltage [V]; N_s Number of cells connected in series.

In this model, R_s models Joule effect Losses, while R_{sh} represents the losses due to the imperfect nature of materials used for solar cells making.

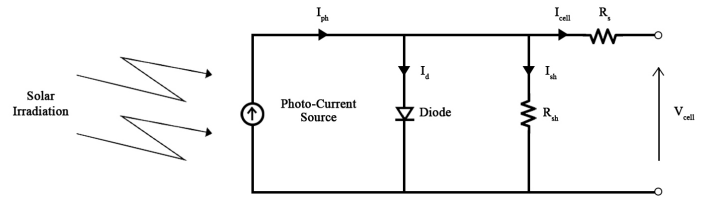


Figure 4: Single diode representation of a PV cell

Several PV cells are connected to obtain a PV array. Connecting them in parallel results in a higher output current, while output voltage increases if they are connected in series. The same thing occurs when passing from PV array to PV fields constituting a PVG. Therefore, depending on the application and the load to be supplied, the PVG should consist of N_s array in series constituting strings and N_p strings connected in parallel. In this study, it is considered an $P_{mpp} = 8.1kW_p$ PVG peak power. The solar array used in the simulation is a 1SOLTECH-1STH-215P from the SIMPOWERSYSTEMS library present in SIMSCAPE in MATLAB/SIMULINK, with a peak power of $P_{array,mpp} = 213.1W_p$ in Standard Test Conditions (STC). It corresponds to a single diode model presented above, whose input parameters are the solar irradiation and the cell temperature. The number of arrays N necessary in order to reach an overall power of about $8.1kW_p$ is equal to 38 according to (6). By choosing N_s and N_p equal to 19 and 2 respectively, the operation range of the PVG voltage will be comprises between $V_{sc}=0V$ and $V_{oc}=689V$, and a DC/DC Boost converter is needed to allow the connection between the PVG output and the DC-bus of 1000V. The $8.1kW_p$ PVG Voltage-Current and Voltage-Power curves, in STC, are presented in Figure 5(a) and Figure 5(b) respectively, depicting the PVG main characteristics.

$$N_{arrays} = \frac{P_{PV}}{P_{array}} = \frac{8100}{215} = 37.34 \tag{6}$$

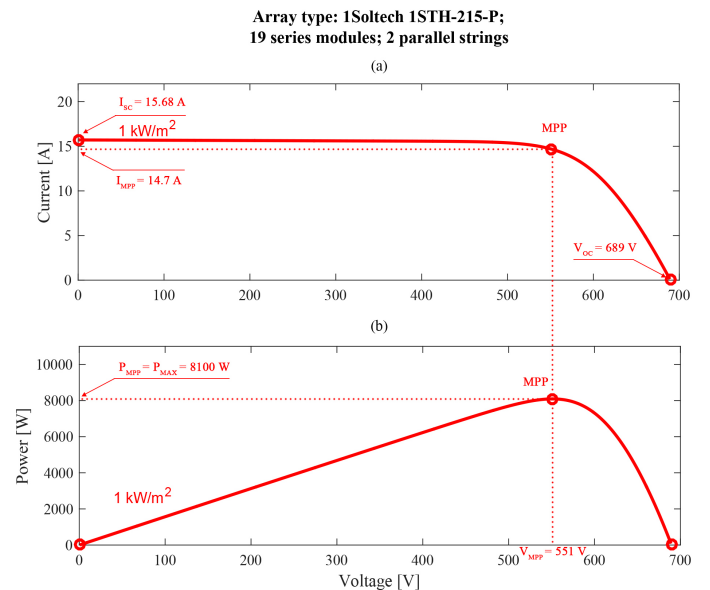


Figure 5: PVG characteristics curves in STC (a) Current-Power (b) Voltage-Power

Equations (7) and (8) show the DC/DC Boost converter input/output voltages and input/output currents relations respectively.

where: $V_{out} = V_{DC}$ DC-bus voltage [V]; $V_{in} = V_{pv}$ PVG output Voltage [V]; I_{out} Boost converter output current [A]; $I_{in} = I_{pv}$ PVG output current [A]; D Duty Cycle of the converter.

By neglecting losses, the relation between the input and output powers of the converter is given by (9). Equations (10) and (11) show the maximum and nominal powers of the DC/DC Boost converter respectively [12]. According to (7), the duty cycle increase causes the PVG voltage decrease, and its decrease results on the PVG voltage increase.

$$V_{out} = V_{in}/(1 - D) \tag{7}$$

$$I_{in} = I_{out}/(1 - D) \tag{8}$$

$$P_{in} = P_{pv} = V_{pv} \cdot I_{pv} = V_{DC} \cdot I_{out} = P_{out} \tag{9}$$

$$P_{Boost}^{max} = P_{PV}^{max} \tag{10}$$

$$P_{Boost}^{nom} = 0.9P_{PV}^{max} \tag{11}$$

2.2. Photovoltaic Generator Control

Depending on the EMS in operation and on the application, the PVG will have to be controlled by two different algorithms. If the fully potential of the PVG is requested, then it will be controlled by the mean of the MPPT, otherwise, if its output power must be limited, the LPPT will take over. In both cases, reaching the operating point is often achieved by introducing a DC-DC Boost Converter, which constitutes a voltage adaptation stage between the PVG output and the DC-Bus. In both cases, the control algorithm generates at its output the duty cycle of the converter, to reach the desired voltage on the PVG voltage-power curve. PWM generates the IGBT switching signals according to this duty cycle. The decision of limiting the power is given by the supervisory system according to the EMS in operation. It is possible to integrate the PVG control via the ASC, but it is not possible in this case as the PVG voltage, varying between $V_{sc}=0V$ and $V_{oc}=689V$, is lower than the 1000V's DC-Bus. The overall PVG architecture control is shown in Figure 6. The characteristics of the DC/DC Boost converter used in the simulation are summarized in Table 1.

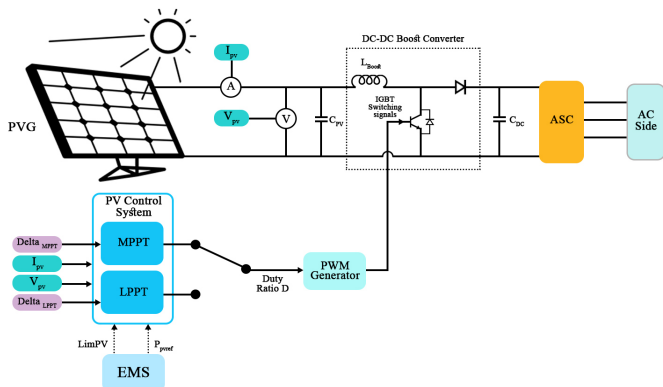


Figure 6: Overall PVG control architecture

By using a P&O control algorithm, the optimal point (V_{MPP}, P_{MPP}) is reached for any value of temperature and solar radiation. The P&O algorithm corresponds to the right part of Figure 9 and its operation on the voltage-power curve is depicted in Figure 7. In case of power limitation, the PVG reference power is calculated by the supervisory system according to the EMS in operation. This power is achieved using the LPPT algorithm. As with the MPPT, the LPPT makes it possible to reach the limited power point (LPP) on the voltage-power curve, by imposing the corresponding voltage on the PVG terminals. Seen the bell shape of this curve, this power corresponds to two voltages. It is preferable to impose the greatest voltage (located on the right side of the MPP) to decrease the PVG output current, in order to decrease the Joule effect losses according to (9). The LPPT step Δ_{LPPT} is reduced when entering the convergence zone (12), in order to reduce the amplitude of the oscillations around the power reference P_{PVref} , and thus to increase tracking precision of this reference. The LPPT algorithm corresponds to the left part of Figure 9 and its operation on the voltage-power curve is depicted in Figure 8.

$$P_{PV} - P_{PVref} < \epsilon \tag{12}$$

Table 1: DC/DC Boost converter characteristics

L_{boost}	0.0001H
P_{Boost}^{nom}	7.29kW
P_{Boost}^{max}	8.10kW

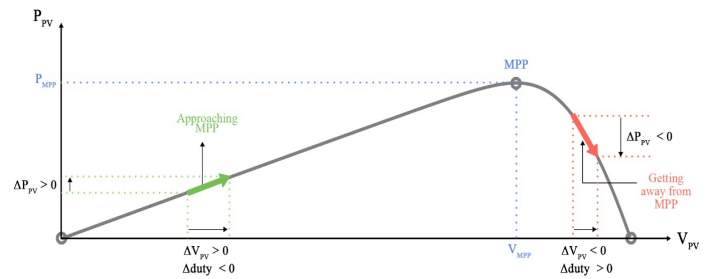


Figure 7: P&O algorithm principle

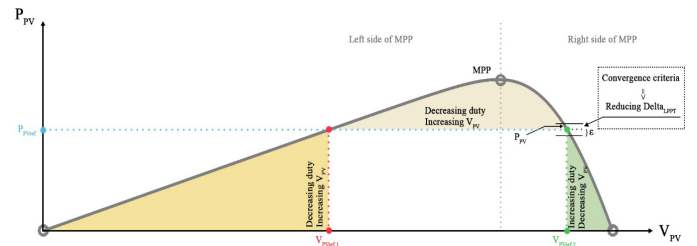


Figure 8: LPPT algorithm principle

2.3. Battery System Mathematical Modelling

Batteries are no longer simple components with reduced number of developed models, so it is not easy to model the electrochemical interactions of a battery by simple circuits. In the literature, many models are available [17]-[20], to meet fine and

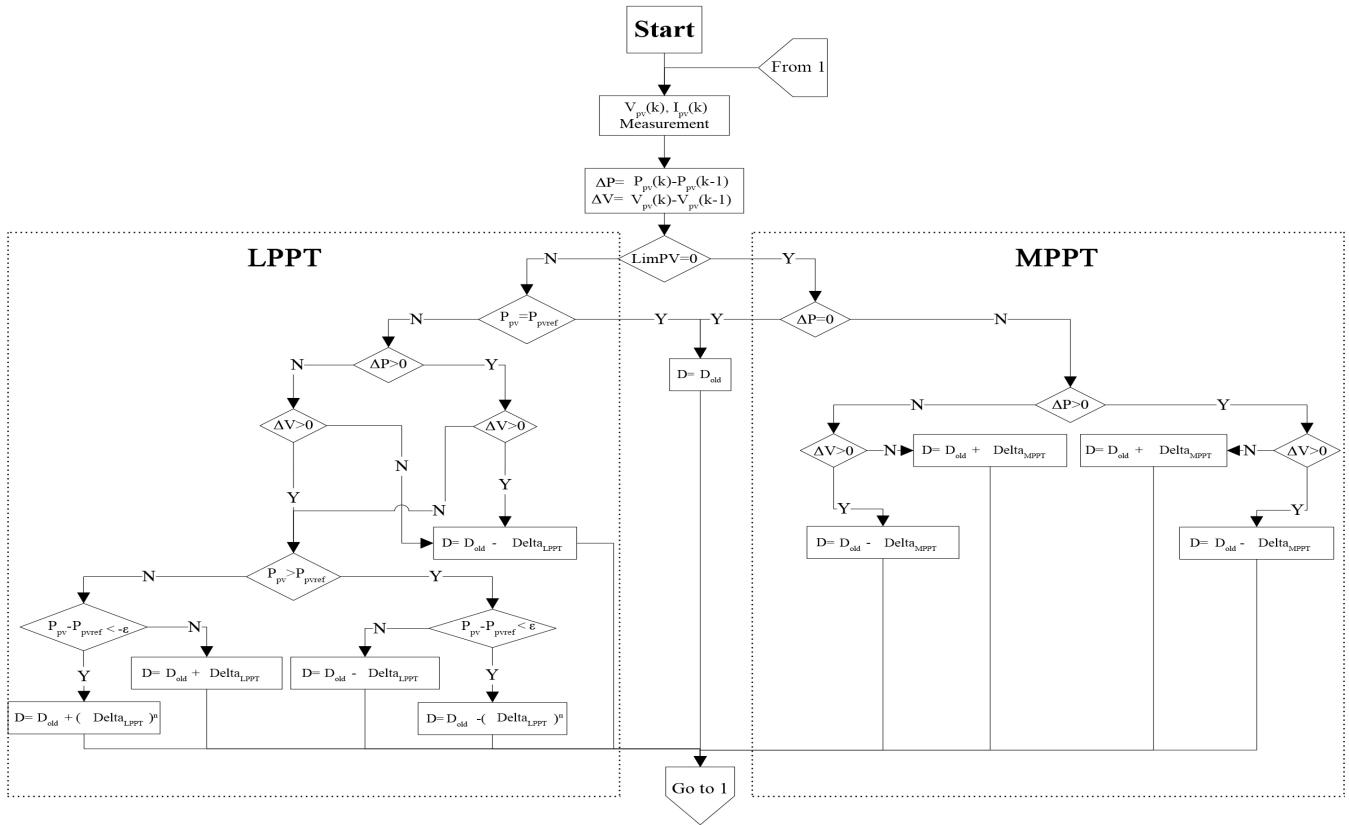


Figure 9: Overall PVG control algorithm

rapid simulation needs. New dynamic models of Lead-Acid batteries are presented in [21], and take into account the dynamic character of the elements of the equivalent circuit, namely the resistance and the capacity of the main branch. In our case, a simple model is sufficient, responding properly to charge and discharge battery requests. In this study’s simulation a Lead-Acid battery is used. It is included in the SIMPOWERSYSTEM Toolbox of MATLAB/SIMULINK. The battery is modeled as a variable voltage source connected in series with an internal resistance, so the voltage across the battery is given by (13). The internal resistance R_{int} of the battery is assumed to be constant during charge and discharge and does not change with the magnitude of the current. Contrary to this, the open circuit voltage E_{bat} is different during charging and discharging processes. It depends on the battery current i_{bat} , on the capacity extracted and on the Hysteresis phenomenon of the battery. Therefore, E_{bat} is given during discharging and charging processes by (14) and (15) respectively, and the instantaneous value of the SOC is given by (16), according to [22].

$$U_{bat} = E_{bat} - R_{int}i_{bat} \quad (13)$$

$$E_{bat-d\acute{e}ch} = E_0 - K \cdot \frac{Q}{Q-i_t} \cdot i^* - K \cdot \frac{Q}{Q-i_t} \cdot i_t + F_{hyst-d\acute{e}ch}(i_{bat}) \quad (14)$$

$$E_{bat-char} = E_0 - K \cdot \frac{Q}{|i_t|+0,1Q} \cdot i^* - K \cdot \frac{Q}{Q-i_t} \cdot i_t + F_{hyst-char}(i_{bat}) \quad (15)$$

$$SOC(\%) = 100 \cdot (1 - \frac{\int i_{bat} \cdot dt}{Q}) \quad (16)$$

where: U_{bat} Voltage at the battery terminals [V]; E_{bat} battery Open-Circuit Voltage [V]; R_{int} battery Internal Resistance [Ω]; i_{bat} battery Current [A]; E_0 Constant Voltage [V]; K Polarization constant or polarization resistance; i^* Dynamic Low Frequency current [A]; i_t Capacity Extracted [A]; Q Maximum battery Capacity [C]; $F_{hyst-char}(i_{bat})$ & $F_{hyst-d\acute{e}ch}(i_{bat})$ battery current Functions representing HYSTERESIS phenomenon of the battery during charge and discharge respectively.

Equations (17) and (18) shows respectively the maximum and nominal powers of the DC/DC Boost converter [12].

$$P_{Buck-Boost}^{max} = \max(|P_{bat_{ch}^{max}}|, P_{bat_{disch}^{max}}) \quad (17)$$

$$P_{Buck-Boost}^{nom} = 0.9P_{Buck-Boost}^{max} \quad (18)$$

where $P_{bat_{ch}^{max}}$ and $P_{bat_{disch}^{max}}$ are the battery maximum charging and discharging powers respectively (battery charging power is counted negatively in MATLAB/SIMULINK). Figure 10 shows the equivalent circuit of the battery according to the model described below. The characteristics of the battery-bank overall system, used in simulation are presented in Table 2, where the complementary informations concerning the battery sizing are presented in section III.

Table 2: Simulation overall battery-bank system characteristics

BBDC Control technique	PWM
Battery Type	Lead Acid
N_s	42
N_p	1

V_{bat}	504V
C_{bat}	100Ah
$P_{bat_{ch}}^{max}$	-4.9kW
$P_{bat_{disch}}^{max}$	4.6kW
L_{BBDC}	0.05H
P_{BBDC}^{nom}	4.41kW
P_{BBDC}^{max}	4.9kW
K_p^{bat}	0.1
K_i^{bat}	10

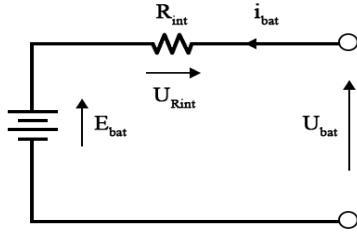


Figure 10: Battery equivalent circuit representation

2.4. Battery Bank Control

The battery-bank is connected to the DC-Bus via a BBDC. The reference power P_{batref} to be delivered or received by the battery-bank is generated by the supervisory according to the EMS in operation. Then, this power is divided by the battery-bank voltage V_{bat} to obtain the current reference I_{bat} . A PI based current loop control is used to adjust the current of the battery-bank at its reference I_{batref} . A PWM generator generates the opposite switching signals S_1 and S_2 of the BBDC, according to the duty cycle D constituting the PI controller output. Figure 11 illustrates this control process.

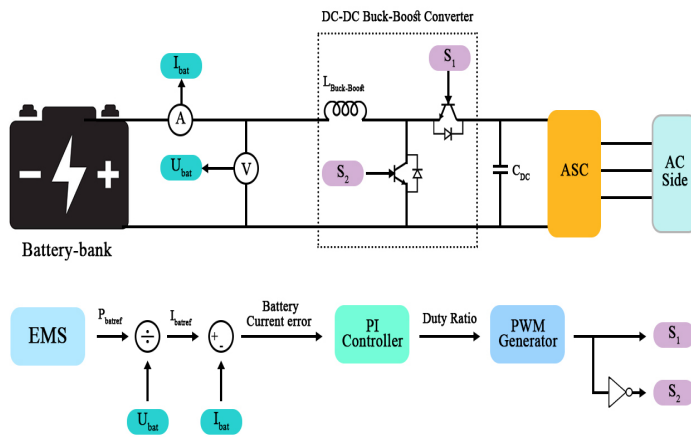


Figure 11: Overall Battery-Bank control architecture

2.5. AC Side Equations

Fig. 12 indicates the alternative side of the GCHRES. Voltage and current equations are given by (19) and (20) respectively.

$$\begin{pmatrix} v_{CCP,a} \\ v_{CCP,b} \\ v_{CCP,c} \end{pmatrix} = \begin{pmatrix} R_F & 0 & 0 \\ 0 & R_F & 0 \\ 0 & 0 & R_F \end{pmatrix} \begin{pmatrix} i_{F,a} \\ i_{F,b} \\ i_{F,c} \end{pmatrix} + \begin{pmatrix} L_F & 0 & 0 \\ 0 & L_F & 0 \\ 0 & 0 & L_F \end{pmatrix} \frac{d}{dt} \begin{pmatrix} i_{F,a} \\ i_{F,b} \\ i_{F,c} \end{pmatrix} + \begin{pmatrix} v_{ASC,a} \\ v_{ASC,b} \\ v_{ASC,c} \end{pmatrix} \quad (19)$$

$$\begin{pmatrix} i_{F,a} \\ i_{F,b} \\ i_{F,c} \end{pmatrix} = \begin{pmatrix} i_{L,a} \\ i_{L,b} \\ i_{L,c} \end{pmatrix} + \begin{pmatrix} i_{UG,a} \\ i_{UG,b} \\ i_{UG,c} \end{pmatrix} \quad (20)$$

where: $j = (a, b, c)$; $i_{F,j}$ currents through the ASC output filter [A]; $i_{L,j}$ load currents [A]; $i_{UG,j}$ UG currents [A]; $v_{CCP,j}$ CCP simple voltages [V]; $v_{ASC,j}$ ASC output voltages [V]; R_F , L_F respectively ASC output Filter resistance [Ω] and inductance [H].

After passing through PARK transformation, the system (19) becomes (21).

$$\begin{pmatrix} v_{CCPd} \\ v_{CCPq} \end{pmatrix} = \begin{pmatrix} R_F & 0 \\ 0 & R_F \end{pmatrix} \begin{pmatrix} i_d \\ i_q \end{pmatrix} + \begin{pmatrix} L_F & 0 \\ 0 & L_F \end{pmatrix} \frac{d}{dt} \begin{pmatrix} i_d \\ i_q \end{pmatrix} + \begin{pmatrix} 0 & \omega L_F \\ -\omega L_F & 0 \end{pmatrix} \begin{pmatrix} i_d \\ i_q \end{pmatrix} + \begin{pmatrix} v_{ASCd} \\ v_{ASCq} \end{pmatrix} \quad (21)$$

where: v_{CCPd}, v_{CCPq} direct and quadrature CCP voltage components respectively [V]; v_{ASCd}, v_{ASCq} direct and quadrature ASC output voltage components respectively [V]; i_d, i_q direct and quadrature filter current components respectively [A]; ω UG Pulsation [rad/s].

2.6. VOC Control of Alternative Side Converter

Controlling the ASC aims to reach two main objectives. The first is to ensure a unity power factor on the AC-side. It means zero reactive power exchange between the ASC and the CCP. The second is to regulate the DC-Bus voltage at its reference value. Keeping its voltage constant is very important since all the power converters are connected to it. The reactive power and DC-Bus voltage references are fixed to $Q_{ccpref} = 0\text{VAR}$ and $V_{DCref} = 1000\text{V}$ respectively. DC-side current equations are given by (22), (23) and (24).

$$i_{HS} = i_{pv} + i_{bat} \quad (22)$$

$$i_{ASCin} = i_{HS} - i_c \quad (23)$$

$$i_c = C_{DC} \frac{dV_{DC}}{dt} = i_{HS} - i_{ASCin} \quad (24)$$

where: i_c DC-Bus Current [A]; i_{ASCin} ASC Input Current [A]; i_{HS} Renewable Hybrid System current [A]; i_{pv} PVG Output Current [A]; i_{bat} Battery charging/discharging current [A]; V_{DC} DC-Bus voltage [V].

Power equations are given by (25) and (26).

$$P_{DC} = V_{DC} \cdot i_c = P_{HS} - P_{ASCin} \quad (25)$$

$$P_{HS} = P_{pv} + P_{bat} \quad (26)$$

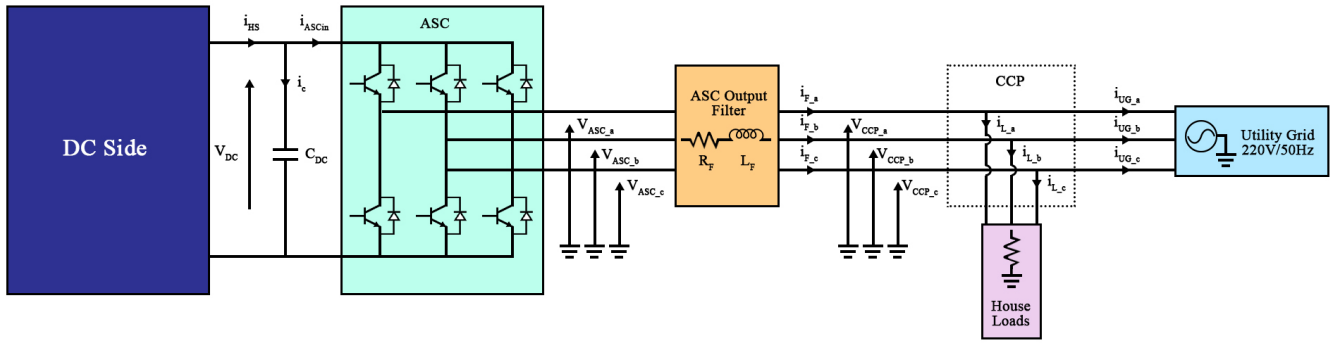


Figure 12: AC-side of the GCHRES

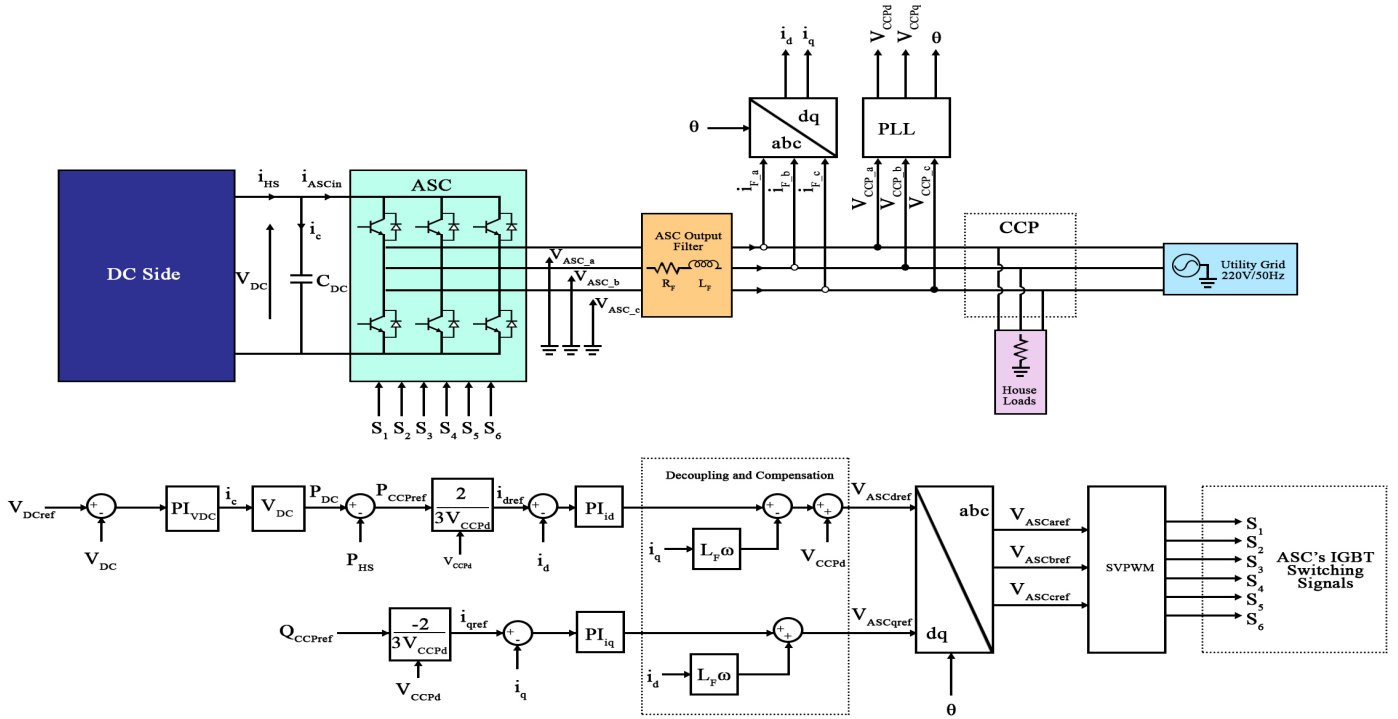


Figure 13: Overall ASC control architecture

where: P_{DC} DC-Bus power [W]; P_{ASCin} ASC input power [W]; P_{HS} renewable hybrid system power [W]; i_{pv} PVG output power [W]; P_{bat} Battery charging/discharging power [W].

From (25), the DC-Bus voltage can be regulated by controlling the ASC input power P_{ASCin} . By neglecting the ASC losses, we can assume that:

$$P_{ASCin} = P_{ASCout} \quad (27)$$

VOC consists on aligning the d-axis of the (d,q) rotational coordinate system with the direct component v_{pccd} of the CCP voltage, thus the CCP voltages will be:

$$v_{CCPd} = V_{CCP} \quad (28)$$

$$v_{CCPq} = 0 \quad (29)$$

The CCP active power is given by (30) and will be simplified to (31) according to (28) and (29).

$$P_{CCP} = \frac{3}{2} (v_{CCPd} \cdot i_d + v_{CCPq} \cdot i_q) \quad (30)$$

$$P_{CCP} = \frac{3}{2} V_{CCP} \cdot i_d \quad (31)$$

And by neglecting the filter resistance, the CCP active power will be:

$$P_{CCP} = P_{ASCout} = P_{ASCin} = \frac{3}{2} V_{CCP} \cdot i_d \quad (32)$$

According to (32), the ASC input power can be controlled by only controlling the d-axis current i_d , since after aligning v_{CCPd} with the d-axis, this latter became constant. Consequently, the DC-Bus power will be controlled too according to (25). The CCP reactive power is expressed by (33) and will also be simplified to (34) according to (28) and (29).

$$Q_{CCP} = \frac{3}{2}(v_{CCPq} \cdot i_d - v_{CCPd} \cdot i_q) \quad (33)$$

$$Q_{CCP} = -\frac{3}{2}V_{CCP} \cdot i_q \quad (34)$$

Therefore, the reactive power exchange between the ASC and the CCP can be controlled by only controlling the q-axis current i_q . According to (28) and (29) the equations system (21) becomes (35), and after passing through Laplace transformation, the system equation obtained in (36) will be used for the ASC control process.

$$\begin{pmatrix} v_{ASCd} \\ v_{ASCq} \end{pmatrix} = \begin{pmatrix} R_F & 0 \\ 0 & R_F \end{pmatrix} \begin{pmatrix} i_d \\ i_q \end{pmatrix} + \begin{pmatrix} L_F & 0 \\ 0 & L_F \end{pmatrix} \frac{d}{dt} \begin{pmatrix} i_d \\ i_q \end{pmatrix} + \begin{pmatrix} 0 & -\omega L_F \\ \omega L_F & 0 \end{pmatrix} \begin{pmatrix} i_d \\ i_q \end{pmatrix} + \begin{pmatrix} V_{CCP} \\ 0 \end{pmatrix} \quad (35)$$

$$\begin{pmatrix} V_{ASCd} \\ V_{ASCq} \end{pmatrix} = \begin{pmatrix} R_F + L_F S & 0 \\ 0 & R_F + L_F S \end{pmatrix} \begin{pmatrix} I_d \\ I_q \end{pmatrix} + \begin{pmatrix} 0 & -\omega L_F \\ \omega L_F & 0 \end{pmatrix} \begin{pmatrix} I_d \\ I_q \end{pmatrix} + \begin{pmatrix} V_{CCP} \\ 0 \end{pmatrix} \quad (36)$$

According to (36), both voltages V_{ASCd} and V_{ASCq} act on the two currents I_d and I_q . In this case the decoupled control technique will be used. By setting a new variable as shown in (37) and (38):

$$E_d = -\omega L_F I_q + V_{CCP} \quad (37)$$

$$E_q = \omega L_F I_d \quad (38)$$

We obtain:

$$\begin{pmatrix} V_{id} \\ V_{iq} \end{pmatrix} = \begin{pmatrix} R_F + L_F S & 0 \\ 0 & R_F + L_F S \end{pmatrix} \begin{pmatrix} I_d \\ I_q \end{pmatrix} + \begin{pmatrix} E_d \\ E_q \end{pmatrix} \quad (39)$$

And by setting a new command variable as shown in (40) and (41):

$$U_d = V_{ASCd} - E_d \quad (40)$$

$$U_q = V_{ASCq} - E_q \quad (41)$$

The currents correctors outputs are obtained as shown in (42) and (43), realizing the decoupled (d,q) axes currents control, as shown in Figure 13, which depict the overall ASC Control scheme.

$$U_{d_{ref}} = P I_d (I_{d_{ref}} - I_d) = (K_p^{Id} + \frac{K_i^{Id}}{s})(I_{d_{ref}} - I_d) \quad (42)$$

$$U_{q_{ref}} = P I_q (I_{q_{ref}} - I_q) = (K_p^{Iq} + \frac{K_i^{Iq}}{s})(I_{q_{ref}} - I_q) \quad (43)$$

A PLL block is used to synchronize the ASC with the UG. Their connection requires to continuously determine the phase angle, on which the control of active and reactive powers mainly depend. PLL is an effective method for determining this angle

The principle is to use a PI corrector to regulate the q-axis CCP voltage component v_{CCPq} to a reference value equal to zero. The output of the PI block is the angular speed of the (d,q) rotational frame reference, and the angle θ is obtained by integrating this

speed as shown in Figure 14. SVPWM control of the ASC is used to reduce current ripples and also the switching frequency of the converter. Equations (44) and (45) shows the maximum and nominal powers of the ASC respectively [12]. Parameters of the AC-side used in simulation are shown in Table 3.

$$P_{ASC}^{max} = P_{bat}^{max} + P_{PV}^{max} \quad (44)$$

$$P_{ASC}^{nom} = 0.9 P_{ASC}^{max} \quad (45)$$

3. Battery Bank Sizing & Energy Management Strategies

3.1. Battery Bank Sizing

Monthly real average solar irradiation and temperature data of the region of Marrakech collected from [23] on hourly basis have been required and used for sizing the battery-bank. These data are shown in Appendix A. The average demand profile estimation for each month, on hourly basis, have been also required. As mentioned above, the battery operates for peak shaving application. Thus, it is sized according to the following criterias: maximum discharging power, maximum charging power, daily energy to be supplied and daily energy to be absorbed. The maximum power that the battery must be able to absorb/deliver are calculated via equations (46) and (47) respectively. The minus sign in these equations are introduced to respect the battery charging and discharging powers signs convention adopted by MATLAB/SIMULINK, where 't' varies on hourly basis while 'j' varies on monthly basis. A more proper sizing would be realized by reducing the basis of the variation of 't' and 'j', for exemple taking a 10min basis for 't' and daily basis for 'j' (j=1 to 365). According to Table B.1 in Appendix B, the battery maximum charging power P_{bat}^{chmax} is around -4.9kW, which corresponds to the most favorable situation throughout the year in terms of difference between PVG generation and load demand. The battery maximum discharging power $P_{bat}^{dischmax}$ is equal to 4.6kW, which corresponds to the worst situation throughout the year according to the same criteria. The maximum daily energy that the battery must be able to absorb and supply are given by (48) and (49) respectively. Sigma (σ) is a coefficient introduced to take into account only moments of the day where the battery must absorbs energy, in the maximum daily energy calculation. Unlike σ , lambda (λ) allows to take into account only the moments of the day where the battery-bank must supply energy, in the minimum daily energy calculation. These coefficients are determined by (50) and (51).

$$P_{bat}^{chmax} = - \max_{j=1..12} (\max_{t=0..24} (P_{PV}^{MPPT}(t) - P_{load}(t) + P_{UG}^{max})) < 0 \quad (46)$$

$$P_{bat}^{dischmax} = - \min_{j=1..12} (\min_{t=0..24} (P_{PV}^{MPPT}(t) - P_{load}(t) + P_{UG}^{max})) > 0 \quad (47)$$

$$E_{bat}^{chmax} = - \sum_{t=1}^{24} \sigma (P_{PV}^{MPPT}(t) - P_{load}(t) + P_{UG}^{max}) < 0 \quad (48)$$

$$E_{bat}^{dischmax} = - \sum_{t=1}^{24} \lambda (P_{PV}^{MPPT}(t) - P_{load}(t) + P_{UG}^{max}) > 0 \quad (49)$$

$$\sigma = 1 \text{ When } P_{PV}^{MPPT}(t) - P_{load}(t) + P_{UG}^{max} > 0 \text{ and } \sigma = 0 \text{ otherwise (50)}$$

$$\lambda = 1 \text{ When } P_{PV}^{MPPT}(t) - P_{load}(t) + P_{UG}^{max} < 0 \text{ and } \lambda = 0 \text{ otherwise (51)}$$

Table 3: Simulation AC-side characteristics

ASC Control technique	SVPWM
R_F	0.3Ω
L_F	0.0054H
P_{ASC}^{nom}	11.43kW
P_{ASC}^{max}	12.7kW
K_p^{vdc}	1.096
K_i^{vdc}	25.35
K_p^{id}	11.7
K_i^{id}	4687.1
K_p^{iq}	11.7
K_i^{iq}	4687.5

Table B.2 in Appendix B, gives the maximum daily energy that the battery must be able to absorb and supply, corresponding to an absorbing energy of -29.983kWh for the yearly most favorable case, and providing energy of 29.649kWh for the yearly worst case. Finally, the Battery bank capacity is given by equation (52). To take advantage of the battery-bank performances without reducing its lifespan, the adopted values of SOC_{min} and SOC_{max} are 20% and 80% respectively, which correspond to the most often recommended values found in the literature of Lead-Acid batteries. Considering a maximum depth of discharge of $DOD_{max} = SOC_{max} - SOC_{min} = 60\%$, and an operating voltage of 504V, the battery-bank capacity is given by (53).

$$C_{bat} = \frac{\max_{j=1...12}(A, B)}{V_{bat} (SOC_{max} - SOC_{min})} \quad (52)$$

where:

$$A = \sum_{t=1}^{24} \sigma(P_{PV}^{MPPT}(t) - P_{load}(t) + P_{UG}^{max})$$

and

$$B = \sum_{t=1}^{24} \lambda(P_{PV}^{MPPT}(t) - P_{load}(t) + P_{UG}^{max})$$

$$C_{bat} = \frac{29983kWh}{504V \cdot 0.6} = 100Ah \quad (53)$$

Hence, 100Ah, 12V battery rating is considered and therefore 42 Batteries are required to connect in series to constitute the battery-bank.

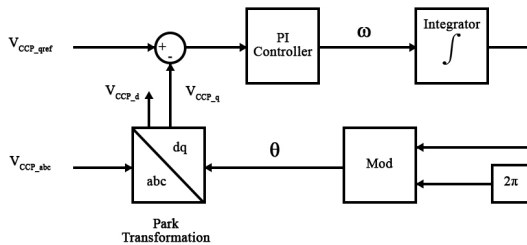


Figure 14: PLL control diagram

3.2. Energy Management Strategies

The different operating modes of the GCHRES and its power flow management are governed by a supervisory controller through EMSs. Three EMSs are proposed in this paper and differ depending on the type of metering with the UG. As mentioned at the introduction, EMS1 is destined for systems with either reversible or irreversible electromechanical metering in the no grid limitation case. EMS2 is destined for the same metering types as EMS 1 but targets the grid limitation case. EMS 3 is destined for digital metering. The main common purpose of all these EMSs is to ensure continuity and reliability of supply to the AC-house, taking into account the battery technical constraints. Hence, they all aim to keep the SOC between a minimum and maximum values SOC_{min} and SOC_{max} respectively. The proper use of the battery avoid reducing its lifespan, avoiding as well the need of replacement of this device. All EMSs aim to charge the battery as soon as possible through the available source; PVG or UG, or both at the same time, to get always operational when requested for peak shaving. The PVG operates by default in MPPT mode, when power limitation is not required. Figure 15 depict the supervisory controller scheme. On the one hand, inputs are related firstly to the type of metering determined by METERTYPE, followed by the UG constraints in term of subscription power P_{UG}^{max} , and maximum injectable power P_{UG}^{injmax} in case of injection limitation (Gridlim=1). On the other hand battery constraints concern the SOC and the maximum charging/discharging powers P_{bat}^{chmax} and $P_{bat}^{dischmax}$ respectively. Power measurement Inputs are the load demand power P_{load} , and the net power P_{net} equal to the difference between PVG and load demand $P_{pv} - P_{load}$. Outputs in terms of battery power reference P_{batref} , and PVG power reference P_{pvref} in case of power limitation (LimPV=1), are generated by the supervisory controller according to the above mentioned Inputs. As mentioned before, EMS 1 and EMS 2 are combined in one flowchart presented in Figure 16, while Figure 17 is representing EMS 3.

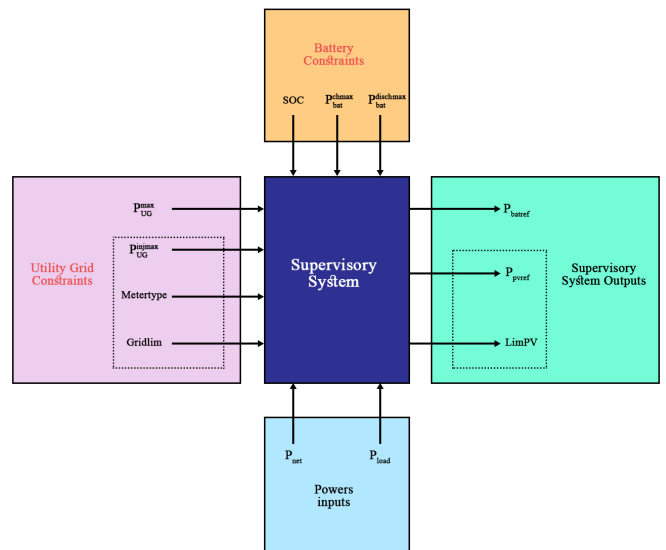


Figure 15: GCHRES supervisory controller scheme

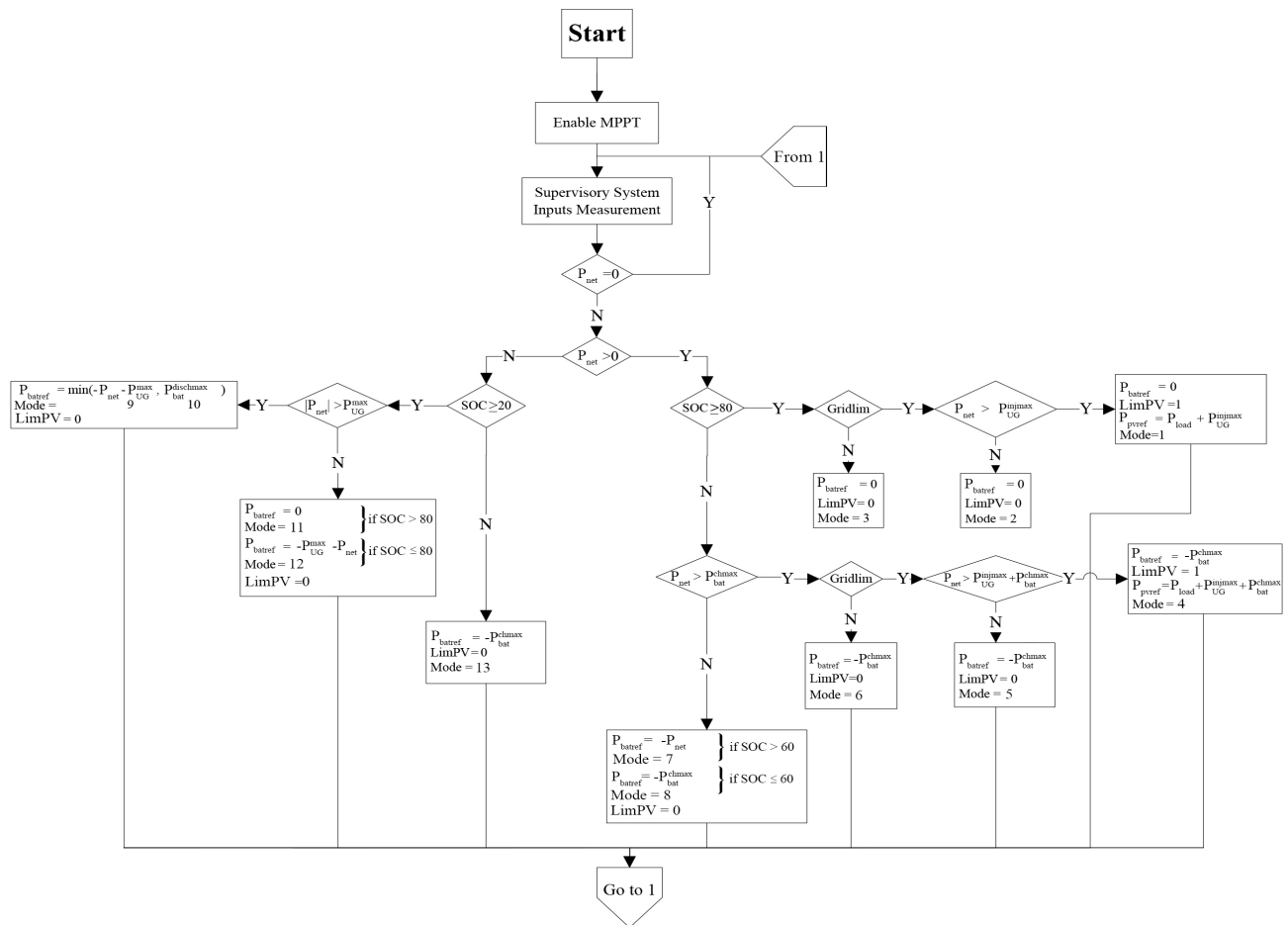


Figure 16: Combined EMS 1 & EMS 2 flowchart

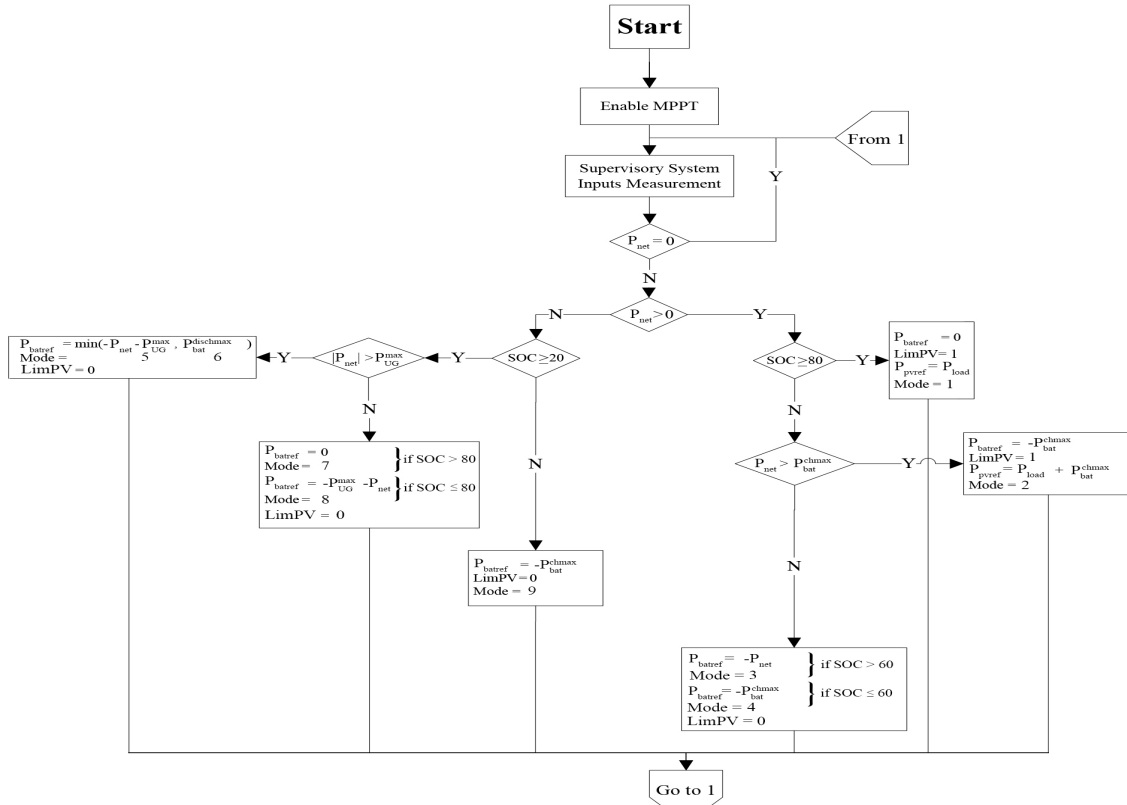


Figure 17: EMS 3 flowchart

4. Simulation Results & Discussion

The GCHRES performances under supervision of each EMS, are tested by simulation on MATLAB/SIMULINK, during a whole day of January. Simulation time are 24 Seconds within the logic of assigning a second to each hour. In order to test the system in all EMSs operating modes, the demand profile P_{load} for this month has been modified. However, the solar irradiation and temperature profiles have not been changed. The UG subscription power is fixed at $P_{UG}^{max} = -5kW$, and the maximum injectable power, in grid injection limitation case is $P_{UG}^{injmax} = 1kW$. Let's remember that the maximum charging power of the battery is $P_{bat}^{chmax} = -4.9kW$, and its maximum discharging power is $P_{bat}^{dischmax} = 4.6kW$. The minimum and maximum battery state of charge SOC fixed at the battery sizing section are respectively $SOC_{min} = 20\%$ and $SOC_{max} = 80\%$. LimPV is the supervisory output that indicates the PVG control algorithm. If $LimPV = 1$ then the PVG is controlled by LPPT, otherwise, if $LimPV = 0$ then the PVG is controlled by MPPT. Note that peak hours run from 18h to 23h in January in Morocco. The adopted sampling time is $T_s = 10^{-5}s$. Each simulation is divided into three time slots, and each time slot is divided into time intervals corresponding to the same operating mode (Opmodes):

- 1st time slot "From $t=0s$ to $t=10s$ ": Solar production is equal to zero $P_{pv}=0$, then begins to increase at $t=8s$, but insufficiently. This phase corresponds to insufficient PVG production ($P_{net}<0$). All EMSs are presenting the same results during this time slot, since their flowcharts are similar when $P_{net}<0$. In order to avoid presenting the same results for many times, only one simulation will be carried out during this time slot, for all EMSs.

- 2nd time slot "From $t=10s$ to $t=18s$ ": Solar production is quite high ($P_{pv} > 0$). Sometimes it exceeds demand ($P_{net}>0$), and that is where the EMSs differ. In effect for the reversible/irreversible electromechanical metering case, if the power is injectable into the UG without limitation, the PVG operates globally in MPPT mode according to EMS 1. For the same metering cases discussed above, if the injectable power into the UG must be limited, PVG operates in LPPT mode in certain cases according to EMS 2. For digital metering case, according to EMS 3, the PVG operates also in LPPT mode in certain situations, to avoid injecting energy into UG. Situations where the PVG control switches toward LPPT are determined by the battery and the UG constraints. A specified demand profile has been dedicated for each EMS for testing their performances in their different operating modes.

- 3rd time slot "From $t=18s$ to $t=24s$ ": The system returns to an operating mode similar to the one on the first time slot, and in which all EMSs have a similar behavior according to their flowcharts. The system presents insufficient production from the PVG ($P_{net}<0$). Consequently, a single simulation will be adopted for all EMSs.

The demand profiles used for simulating these EMSs at the different time slots are presented in Figure 18.

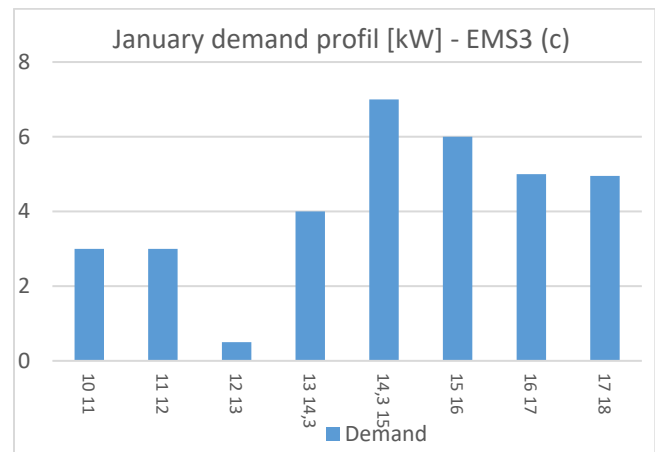
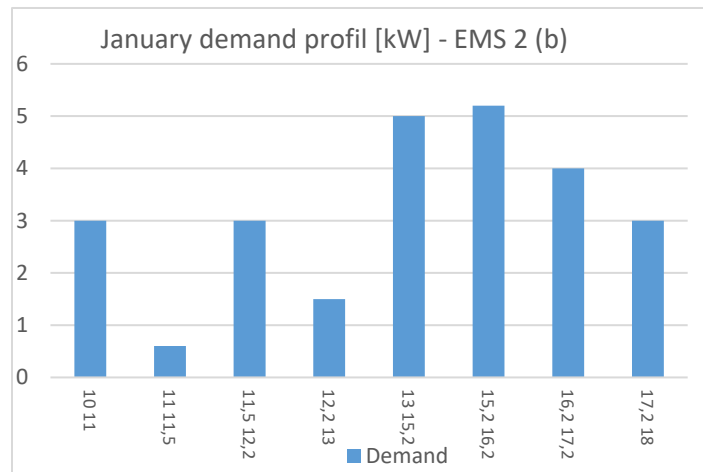
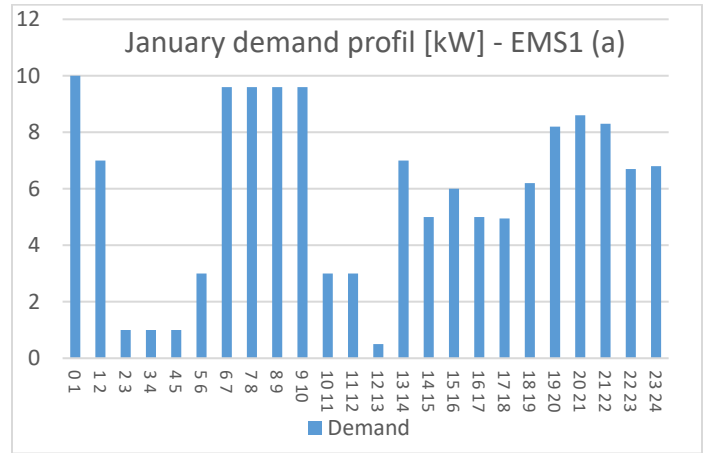


Figure18: EMSs January demand profiles (a) EMS 1 whole day (b) EMS 2 between 10sec and 18sec (c) EMS 3 between 10sec and 18sec

4.1. Energy Management Strategy 1 (EMS1)

1st Time slot "From $t=0s$ to $t=10s$ "

The operating modes (OPmode) and the PVG limitation power decision (LimPV) are presented in Figure 19. Figure 20 presents the battery SOC evolution during this time slots. The net power (P_{net}) is presented in Figure 21, and the different system powers are all depicted in Figure 22.

Figure 23 shows the battery power (P_{bat}) with its reference P_{batref} , and shows also the respect of the battery power constraints P_{bat}^{chmax} and $P_{bat}^{dischmax}$. UG power (P_{UG}), respecting its constraint P_{UG}^{max} , is illustrated in Figure 24. It will be assumed that the battery has an initial SOC of 75%, due to its operation during the previous 24 hours, and in the aim to make the simulation as realistic as possible. According to Figure 20, the battery was sufficiently charged throughout this time slot ($SOC_{t=0s} = 75\%$ and $SOC_{t=10s} = 54.65\%$). Thus, the SOC constraint was respected, and therefore the battery had been operational throughout this time slots. From Figure 21, $P_{net} < 0$ throughout this time slot, and hence the system is on solar generation deficit. The PVG had worked under MPPT control throughout this simulation.

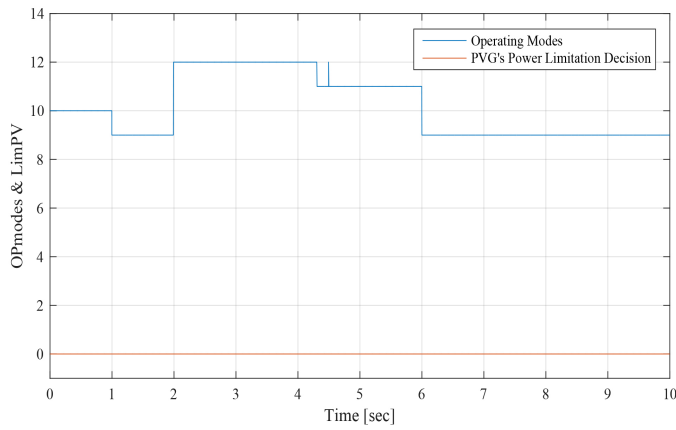


Figure 19: Operating modes and PVG limitation decision between 0sec and 10sec (EMS 1)

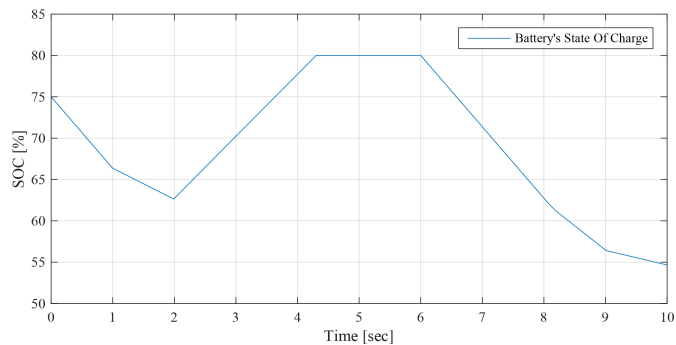


Figure 20: Battery SOC evolution between 0sec and 10sec (EMS 1)

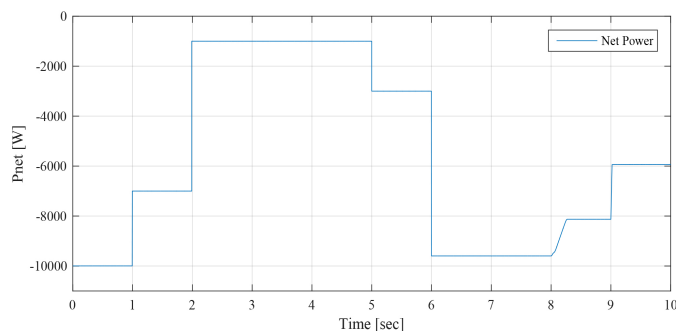


Figure 21: Solar net power between 0sec and 10sec (EMS 1)

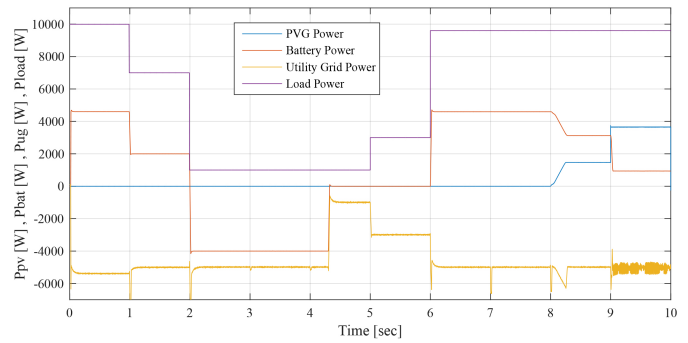


Figure 22: PVG, battery, UG and load powers between 0sec and 10sec (EMS 1)

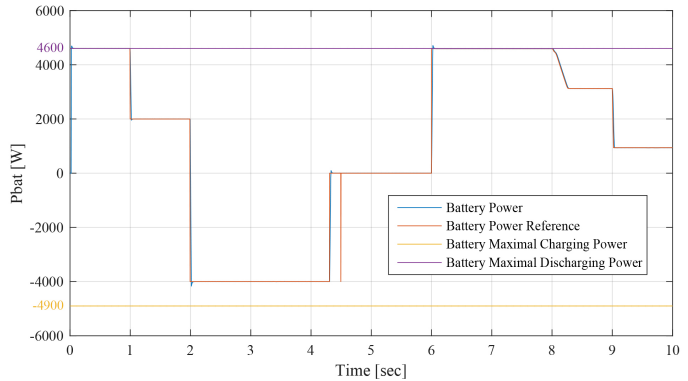


Figure 23: Battery and battery reference powers between 0sec and 10sec (EMS 1)

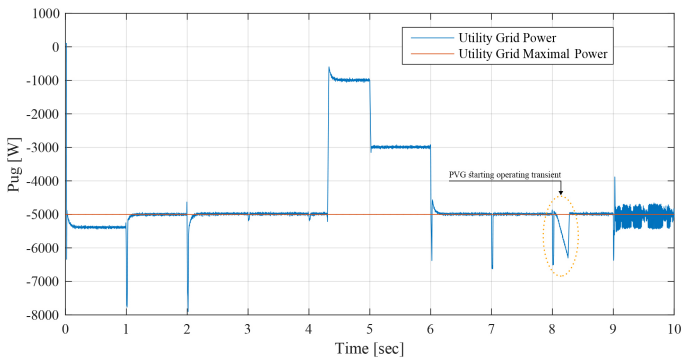


Figure 24: UG power between 0sec and 10sec (EMS 1)

$Op_{mode} = 10 \rightarrow$ from $t=0s$ to $t=1s$: demand is excessively high ($P_{load} = 10kW$) and slightly exceeds the power given by associating the UG with the battery at their maximum operating powers ($P_{bat}^{dischmax} + |P_{UG}^{max}| = 4.6kW + |-5kW| = 9.6kW < |P_{net}|$). The battery performs its peak shaving function while respecting its constraint by not exceeding its maximum discharge power ($P_{batref} = P_{bat}^{dischmax}$). The aim of this EMS is to continuously meet demand, while respecting the technical constraints related to the battery. Therefore, the UG is forced to slightly exceed its maximum power with a value of $-0.4kW$ ($P_{UG} = P_{UG}^{max} - 0.4kW = -5kW - 0.4kW = -5.4kW$), in order to preserve internal circuits of the battery. This will rarely occur since in reality, late at night, demand is quite low. But as said previously, the purpose of modifying the demand profile is to test the performances of the GCHRES during the different EMS operation. This little overflow, which occurs rarely, will be acceptable since in return, the battery, considered as a highly

vulnerable device, will be protected, and thus its lifespan will be extended. The SOC decreases rapidly as the battery discharges with its maximum power.

Opmode = 9 → from t=1s to t=2s & from t=6s to t=10s: On this time intervals, PVG/UG/battery combination is able to satisfy demand ($P_{pv} = 0$ before t=8s thus $P_{net} < 0$). Before t=8s, demand exceeds the UG maximum Power P_{UG}^{max} and the latter constitutes the primary power source ($P_{UG} = P_{UG}^{max}$), while the battery performs its peak shaving function. Its SOC decreases in function of the requested power from the battery ($P_{batref} = P_{load} + P_{UG}^{max} > 0$), which is lower than its maximum value $P_{bat}^{dischmax}$. Around t=8s, sun begins to rise, and solar energy becomes the primary energy source destined to meet demand, but remains insufficient ($P_{net} < 0$), since solar irradiation is too low at the first Hours following sunrise. The UG brings the deficit while not exceeding its maximum power ($P_{UG} = P_{UG}^{max}$), then the battery is discharged in function of PVG+UG deficit ($P_{batref} = |P_{net}| + P_{UG}^{max} > 0$), performing its peak shaving application. During this phases, the power requested by the battery is lower than previous phase (between t=0s and t=1s), which explains the SOC low decrease. A brief exceeding of the maximum power of the UG is noticed from t=8s to t=8.3s due to the MPPT control transient, which requires a little time for tracking the MPP, given that $P_{pv} = 0$ before t=8s, hence the UG compensate this transient. It should be noticed that the transient lasts about 0.3, which is totally negligible in the real life operation of the system throughout the 24 hours of the day. Since it starts generating power, PVG operates in MPPT mode.

Opmode = 12 → from t=2s to t=4.3s: Demand is quite low, and UG can satisfy it alone without reaching its maximum power P_{UG}^{max} . As mentioned previously, all EMSs aim to charge the battery as soon as possible with the available source to get operational once peak shaving is requested. Therefore, the UG satisfies the demand, and the battery is charged by the difference between the maximum UG power and the demand ($P_{batref} = P_{UG}^{max} + P_{load} < 0$). The battery charging power does not exceed its maximal value P_{bat}^{chmax} . In this time period, the UG power is always equal to its maximum ($P_{UG} = P_{UG}^{max}$). SOC increases according to the value of ($P_{UG}^{max} + P_{load}$). At t=4.3s, the SOC reached its maximum (SOC_{max}), and the charging operation got stopped ($P_{batref} = 0$).

Opmode = 11 → from t=4.3s to t=6s: Demand does not exceed the UG maximum power ($P_{load} < |P_{UG}^{max}|$). Since the battery SOC reaches its maximum SOC_{max} , the charging process is no longer authorized ($P_{batref} = 0$). As a result, the UG is strained only to meet demand ($P_{UG} = -P_{load}$). The SOC remains stable at the value of 80%, and the battery power is equal to zero ($P_{batref} = 0$).

2nd Time slot "From t=10s to t=18s"

The operating modes (OPmode) and the PVG limitation power decision (LimPV) are presented in Figure 25. Since the injection into UG is not limited, the LPPT control is never activated (LimPV=0). Therefore, the PVG operates continuously in MPPT

mode. Figure 26 presents the battery SOC evolution. The net power (P_{net}) is presented in Figure 27, and the different system powers are depicted together in Figure 28. Figure 29 shows the battery power (P_{bat}) with its reference (P_{batref}), and shows also the respect of the battery power constraints P_{bat}^{chmax} and $P_{bat}^{dischmax}$. In the beginning of this time slots, the battery $SOC_{t=10s} = 54.65\%$. From Figure 27, $P_{net} > 0$ between t=10s and t=15s, and $P_{net} < 0$ between t=15s and t=18s. Therefore, the system begins with a surplus of solar generation and then gets into deficit situation.

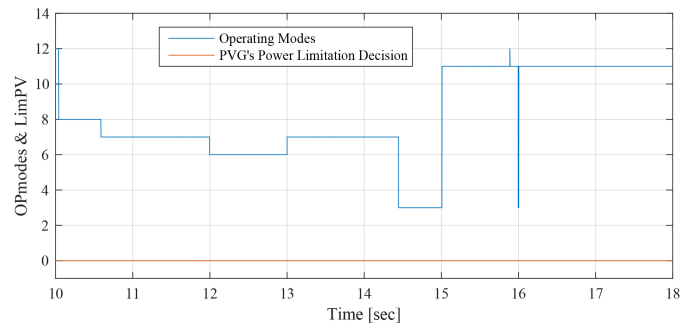


Figure 25: Operating modes and PVG limitation decision between 10sec and 18sec (EMS 1)

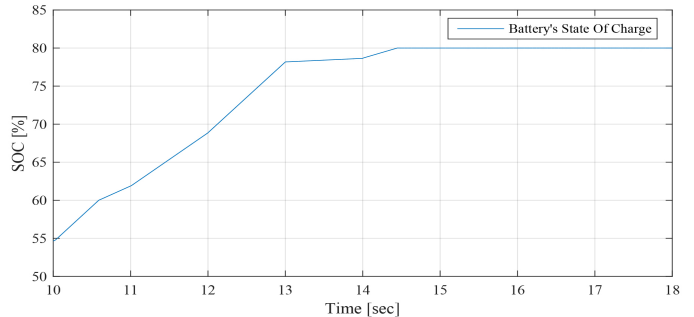


Figure 26: Battery SOC evolution between 10sec and 18sec (EMS 1)

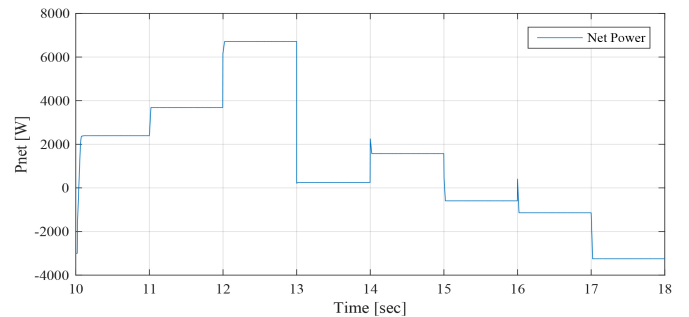


Figure 27: Solar net power between 10sec and 18sec (EMS 1)

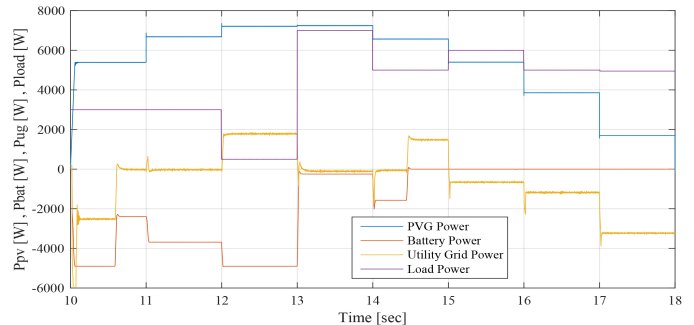


Figure 28: PVG, battery, UG and load powers between 10sec and 18sec (EMS 1)

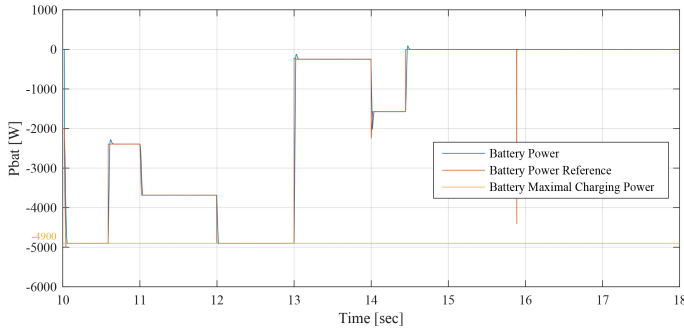


Figure 29: Battery and battery reference powers between 10sec and 18sec (EMS 1)

Opmode = 8 → from t=10s to t=10.6s: PVG power increases with solar irradiation increase. It is sufficient to satisfy the load demand but still less than battery maximum charging power ($0 < P_{net} < |P_{bat}^{chmax}|$). This surplus is used to charge the battery, since its SOC is lower than its maximum value (SOC_{max}). And in order to ensure rapid charging of the battery (due to the SOC value inferior to 60%), aiming to obtain sufficient SOC for the next peak shaving request, the UG provides the necessary power, added to the surplus ($P_{UG} = P_{bat}^{chmax} + P_{net} < 0$), in order to charge the battery with its maximum power ($P_{batref} = P_{bat}^{chmax}$). Therefore, it is noticed that the SOC increases rapidly during this period.

Opmode = 7 → from t=10.6s to t=12s & From t=13s to t=14.3s: PVG is largely sufficient to satisfy the demand, but the surplus is still less than the battery maximum charging power ($0 < P_{net} < |P_{bat}^{chmax}|$). This surplus is used to charge the battery, since its SOC is lower than its maximum value (SOC_{max}). The latter being quite high ($SOC > 60\%$), the UG power is not requested ($P_{UG} = 0$), and the battery is charged only by the solar generation surplus ($P_{batref} = -P_{net}$), which does not exceed the maximum charging power of the battery. Thus, its SOC increases, and its evolution depends on the value of $-P_{net}$. Since SOC reaches its maximum, the battery charging is no longer authorized ($P_{batref} = 0$ at t = 14.3s).

Opmode = 6 → from t=12s to t=13s: PVG is largely sufficient to satisfy the demand and the surplus is greater than the battery maximum charging power ($P_{net} > |P_{bat}^{chmax}|$). This is used to charge the battery with its maximum charging power ($P_{batref} = P_{bat}^{chmax}$), since its SOC is lower than its maximum limit SOC_{max} . The remaining power is injected into the UG ($P_{UG} = P_{net} + P_{bat}^{chmax} > 0$).

Opmode = 3 → from t=14.3s to t=15s: PVG is largely sufficient to satisfy the demand ($P_{net} > 0$). As the battery had reached its maximum SOC_{max} , charging operation is no longer authorized ($P_{batref} = 0$). The battery SOC remains stable at its maximum. The total solar generation surplus is injected into the UG ($P_{UG} = P_{net}$).

Opmode = 11 → from t=15s to t=18s: The system return in the operating mode 11. Demand exceeds PVG production ($P_{net} < 0$). The deficit being less than the UG maximum power ($|P_{net}| < |P_{UG}^{max}|$), only the latter is requested to supports the PVG to meet

demand ($P_{UG} = P_{net} < 0$). However, the battery which had reached its maximum state of charge SOC_{max} at t=15s, remains at rest ($P_{batref} = 0$) and its SOC stable at its maximum recommended.

3rd Time slot “From t=18s to t=24s”

The operating modes (OPmode) and the PVG limitation power decision (LimPV) are presented in Figure 30. Figure 31 present the battery SOC Evolution during this time slots. The net power P_{net} is presented in Figure 32, and the different system power are depicted together in Figure 33. Figure 34 shows the battery power P_{bat} with its reference P_{batref} , and shows also the respect of the battery power constraints P_{bat}^{chmax} and $P_{bat}^{dischmax}$. UG power P_{UG} , respecting its constraint P_{UG}^{max} , is illustrated in Figure 35. In the beginning of this time slots, the battery SOC is maximum recommended (SOC_{max}). At the evening, the PVG stops producing electricity and all demand will have to be met by the UG grid first, backed up by the battery performing the peak shaving. In effect, the UG contributes with its maximum power ($P_{UG} = P_{UG}^{max}$), since the demand exceeds it ($P_{load} > |P_{UG}^{max}|$). The battery provides the deficit (peak shaving), which is less than its maximum discharging power $P_{bat}^{dischmax}$, according to the difference between the maximum UG power and the demand ($P_{batref} = P_{UG}^{max} + P_{load}$). This lead to reduce the call of the UG during this time slot corresponding to peak hour interval in Morocco, reducing therefore the risks of grid congestion. The SOC is consequently in permanent decrease, reaching at the end of the day a value of 53%.

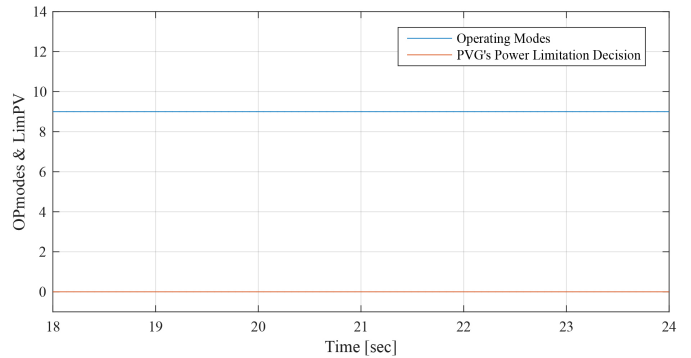


Figure 30: Operating modes and PVG limitation decision between 18sec and 24sec (EMS 1)

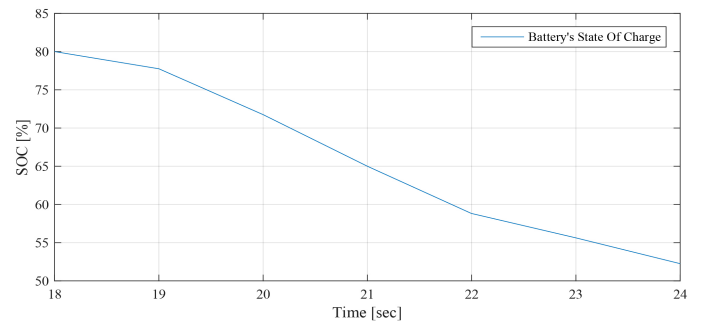


Figure 31: Battery SOC evolution decision between 18sec and 24sec (EMS 1)

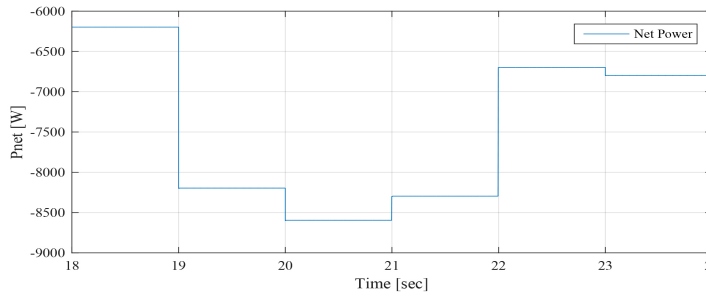


Figure 32: Solar net power between 18sec and 24sec (EMS 1)

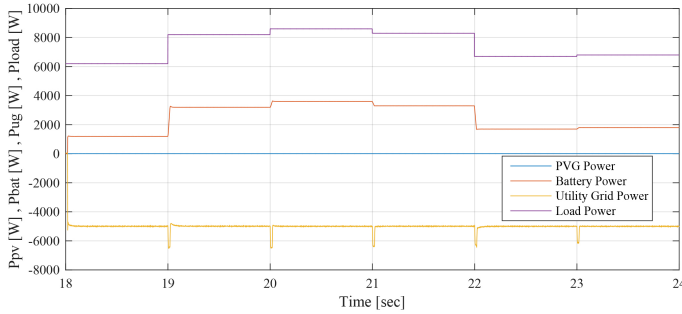


Figure 33: PVG, battery, UG and load powers between 18sec and 24sec (EMS1)

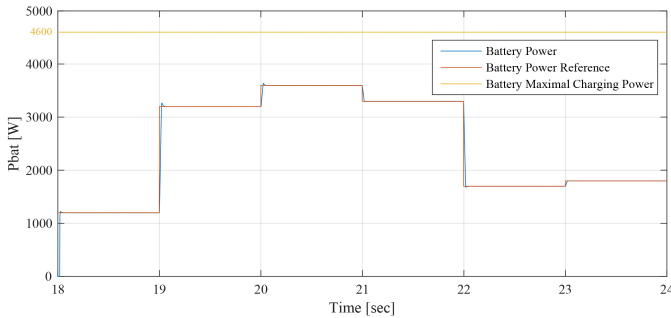


Figure 34: Battery and battery reference powers between 18sec and 24sec (EMS1)

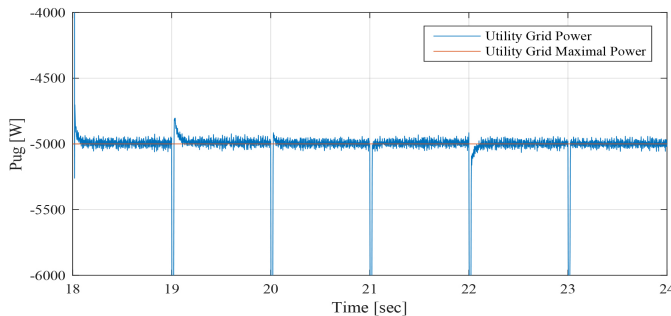


Figure 35: UG powers between 18sec and 24sec (EMS 1)

4.2. Energy Management Strategy 2 (EMS2) 2nd Time slot "From t=10s to t=18s"

As mentioned before, 1st and 3rd time slots is simulated only once and it represent the performances of the GCHRES supervised by each one of the three EMSs. Therefore, only the 2nd time slot is simulated for EMS 2, and starts with the same conditions as those of the EMS 1. The operating modes (OPmode) and the PVG limitation power decision (LimPV) are presented in Figure 36. Figure 37 presents the battery SOC Evolution during this time slots. The net power (P_{net}) is presented in Figure 38, and the different system powers are depicted together in Figure 39.

Figure 40 shows the battery power (P_{bat}), with its reference P_{batref} , and shows also the respect of the battery power constraints P_{bat}^{chmax} and $P_{bat}^{dischmax}$. The UG power, respecting its constraint P_{UG}^{injmax} , is illustrated in Figure 41. In this figure, it is shown that in case of injection into the UG, the power never exceeded 1kW, achieving consequently the purpose of keeping the UG voltage, below the maximum admissible voltage. Figure 42 shows the PVG output power (P_{pv}), and its reference P_{pvref} (when operating in LPPT). The PVG operating voltages (V_{pv}) are presented in Figure 43. From Figure 38, it is clear that $P_{net} > 0$ from t=10s to t=16s, and $P_{net} < 0$ from t=16s to t=18s. Therefore, the system begins with a surplus of solar generation and then gets into deficit situation.

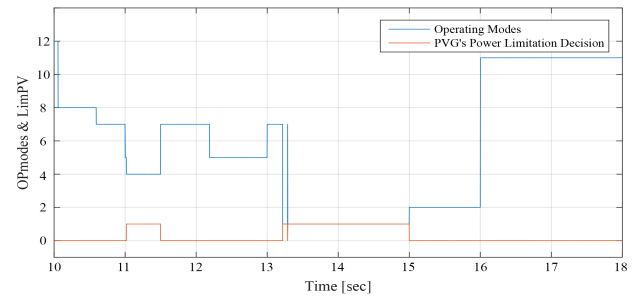


Figure 36: Operating modes and PVG limitation decision between 10sec and 18sec (EMS 2)

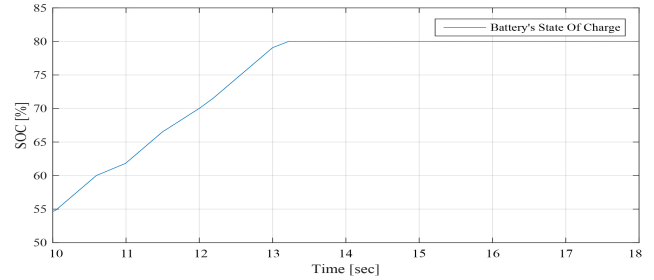


Figure 37: Battery SOC evolution between 10sec and 18sec (EMS 2)

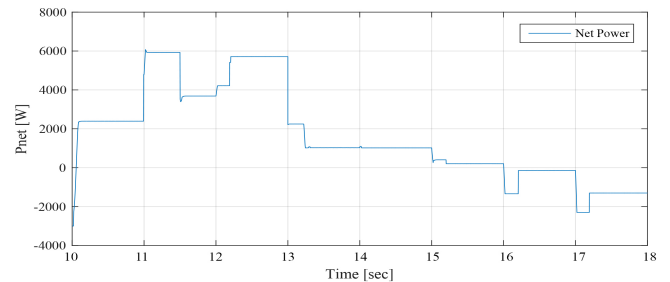


Figure 38: Solar net power between 10sec and 18sec (EMS 2)

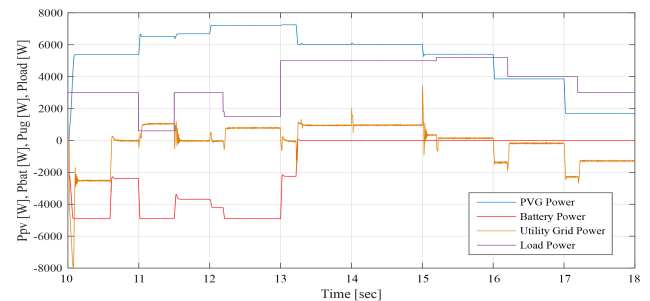


Figure 39: PVG, battery, UG and load powers between 10sec and 18sec (EMS 2)

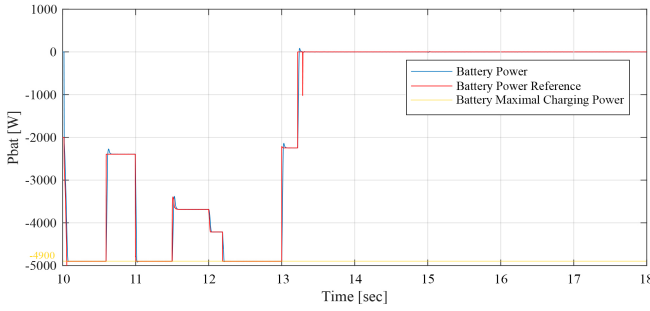


Figure 40: Battery and battery reference powers between 10sec and 18sec (EMS2)

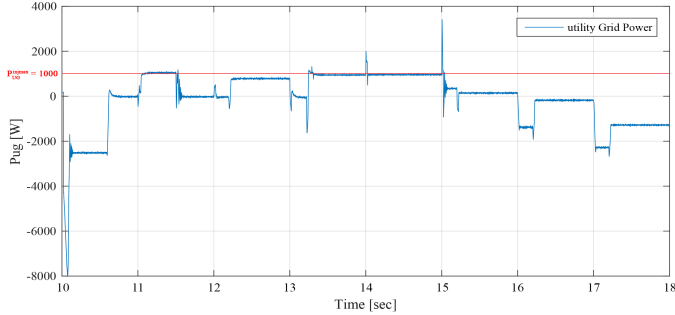


Figure 41: UG power between 10sec and 18sec (EMS 2)

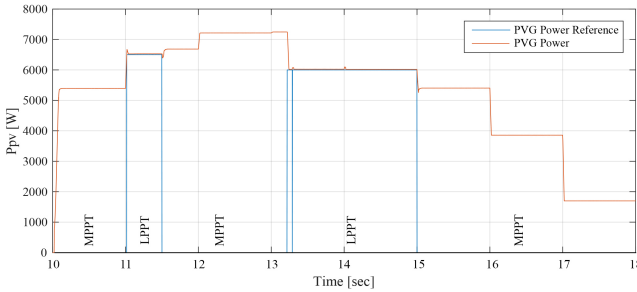


Figure 42: PVG and PVG reference powers between 10sec and 18sec (EMS 2)

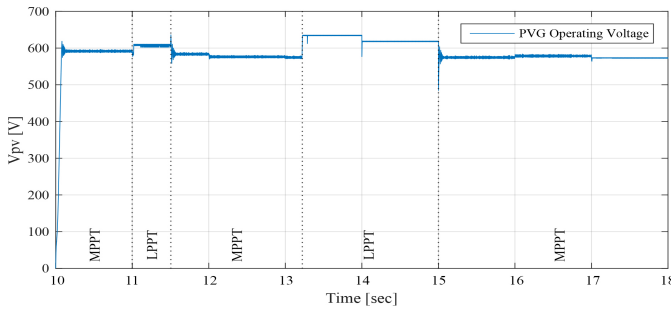


Figure 43: PVG operating voltage between 10sec and 18sec (EMS 2)

Opmode = 8 → from t=10s to t=10.6s: PVG power increases with solar irradiation increase. It is sufficient to satisfy the load demand but still less than battery maximum charging power ($0 < P_{net} < |P_{bat}^{chmax}|$). Hence, the PVG operates in MPPT mode. This surplus is used to charge the battery, since its SOC is lower than its maximum value (SOC_{max}). And in order to ensure rapid charging of the battery (due to the SOC value inferior to 60%) aiming to obtain sufficient SOC for the next peak shaving request, the UG provides the necessary power, added to the surplus ($P_{UG} = P_{bat}^{chmax} + P_{net} < 0$), in order to charge the battery with

its maximum power ($P_{batref} = P_{bat}^{chmax}$). Therefore, it is noticed that the SOC increases rapidly during this period.

Opmode = 7 → from t=10.6s to t=11s & from t=11.5s to t=12.2s & from t=13s to t=13.2s: PVG is largely sufficient to satisfy the demand, but the surplus is still less than the battery maximum charging power ($0 < P_{net} < |P_{bat}^{chmax}|$). Hence, the PVG operates under MPPT control. This surplus is used to charge the battery, since its SOC is lower than its maximum value (SOC_{max}). The latter being quite high ($SOC > 60\%$), the UG power is not requested ($P_{UG} = 0$), and the battery is charged only by the solar generation surplus ($P_{batref} = -P_{net}$), which does not exceed the maximum charging power of the battery. Thus, the SOC increases and its evolution depends on the value of $-P_{net}$. Since SOC reaches its maximum recommended, the battery charging is no longer authorized ($P_{batref} = 0$ at t = 13.2s).

Opmode = 4 → from t=11s to t=11.5s: PVG is largely sufficient to satisfy demand, and the surplus $P_{net} = 6kW$ is higher than the battery maximum charging power added to the UG maximum power injectable ($|P_{bat}^{chmax}| + P_{UG}^{injmax} = |-4.9kW| + 1kW = 5.9kW$). Therefore, The LimPV supervisory output takes the value 1 in order to switch to the LPPT control of the PVG, whose output power P_{pv} follows the reference P_{pvref} . The PVG power is limited to the value $P_{pvref} = P_{load} + |P_{bat}^{chmax}| + P_{UG}^{injmax} = 0.6kW + 4.9kW + 1kW = 6.5kW$, and is reached by imposing the highest of the two voltages making it possible to reach this operating point, located in the right side of the MPP voltage on the (v_{pv}, P_{pv}) Curve. The SOC being less than its maximum value SOC_{max} , the battery is charged via its maximum charging power ($P_{batref} = P_{bat}^{chmax}$) and a power of 1kW is injected into the UG ($P_{UG} = P_{UG}^{injmax}$).

Opmode = 5 → from t=12.2 to t=13 s: PVG is largely sufficient to satisfy demand. Surplus ($P_{net} = 5.71kW$) is higher than battery maximum charging power but remain lower than the latter added to the UG maximum injectable power ($P_{UG}^{injmax} + |P_{bat}^{chmax}| = 1kW + |-4.9kW| = 5.9kW$). Therefore, the PVG operates under the control of the MPPT algorithm. The battery, whose SOC has not yet reached its maximum SOC_{max} , charges via its maximum power ($P_{batref} = P_{bat}^{chmax}$), and the remaining power ($P_{UG} = P_{net} - |P_{bat}^{chmax}| = 5.71kW - 4.9kW = 0.81kW$) is injected into the UG.

Opmode = 1 → from t=13.2s to t=15s: PVG is highly sufficient to satisfy demand ($P_{net} > 0$). Battery charging is no longer authorized ($P_{batref} = 0$) since its SOC had reached its maximum value SOC_{max} at t=13.2s. As the injection into the grid is limited, The LimPV output took the value 1 to switch toward LPPT control. The PVG output power is limited to the value of $P_{pvref} = P_{load} + P_{UG}^{injmax} = 5kW + 1kW = 6kW$, generated by the supervisory system, and a 1kW power was injected into the UG ($P_{UG} = P_{UG}^{injmax}$). SOC value is stable at its maximum value SOC_{max} . As for time interval starting from t=11s and finishing at t=11.5s, the voltage imposed by the LPPT algorithm is the one higher than the MPP voltage.

Opmode = 2 → from t=15s to t=16s: PVG is highly sufficient to satisfy demand ($P_{net} > 0$). As the battery charging is no longer allowed ($P_{batref} = 0$), the surplus is totally injected into the UG ($P_{UG} = P_{net}$), since the maximum injectable power P_{UG}^{injmax} is higher than the solar generation surplus ($P_{UG}^{injmax} > P_{net}$). The PVG is to be controlled by MPPT.

Opmode = 11 → from t=16s to t=18s: Demand exceeds PVG production ($P_{net} < 0$) and therefore the PVG is controlled by MPPT. The deficit being less than the UG maximum power ($|P_{net}| < |P_{UG}^{max}|$), only the latter is requested to supports the PVG to meet demand ($P_{UG} = P_{net} < 0$). However, the battery which had reached its maximum state of charge SOC_{max} at t=15s, remains in rest ($P_{batref} = 0$) and its SOC stable at its maximum.

4.3. Energy Management Strategy 3 (EMS3) 2nd Time slot "From t=10s to t=18s"

The simulation starts with the same conditions as those of the EMS 1 and EMS 2 ($SOC_{t=10s} = 54.65\%$). The operating modes (OPmode) and the PVG limitation power decision (LimPV) are presented in Figure 44. Figure 45 presents the battery SOC Evolution during this time slots. The net power (P_{net}) is presented in Figure 46, and the different system powers are depicted together in Figure 47. Figure 48 shows the battery power (P_{bat}), with its reference P_{batref} , and shows also the respect of the battery power constraints P_{bat}^{chmax} and $P_{bat}^{dischmax}$. The UG power, respecting its constraint P_{UG}^{max} , is illustrated in Figure 49. It proves also that zero power was injected into the UG, avoiding thereby the occurring of an absurd increase of the subscriber energy bill. Figure 50 shows the PVG output power (P_{pv}), and its reference P_{pvref} (when operating in LPPT mode). The PVG operating voltages (V_{pv}) are presented in Figure 51. From Figure 46, it is clear that $P_{net} > 0$ from t=10s to t=14.3s, and $P_{net} < 0$ from t=14.3s to t=18s. Therefore, the system begin with a surplus of solar generation and then gets into deficit situation.

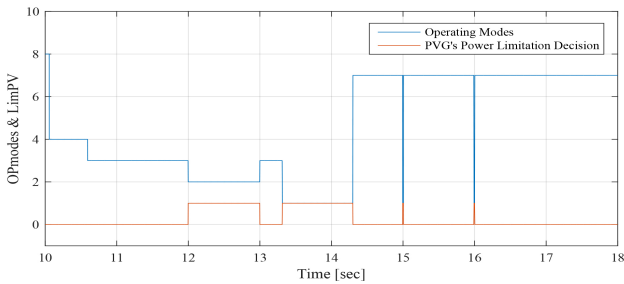


Figure 44: Operating modes and PVG limitation decision between 10sec and 18sec (EMS 3)

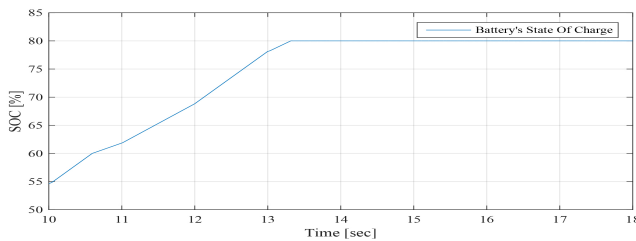


Figure 45: Battery SOC evolution between 10sec and 18sec (EMS 3)

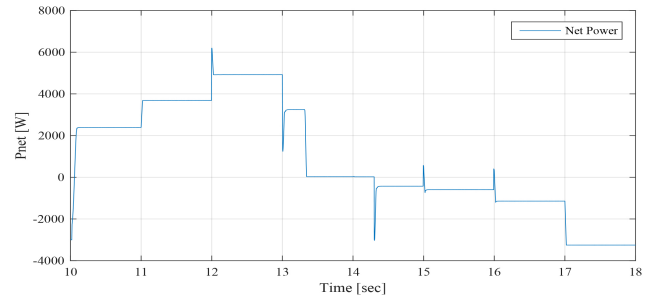


Figure 46: Solar net power between 10sec and 18sec (EMS 3)

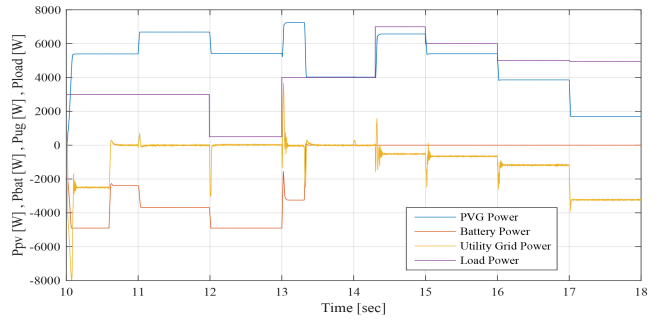


Figure 47: PVG, battery, UG and load powers between 10sec and 18sec (EMS 3)

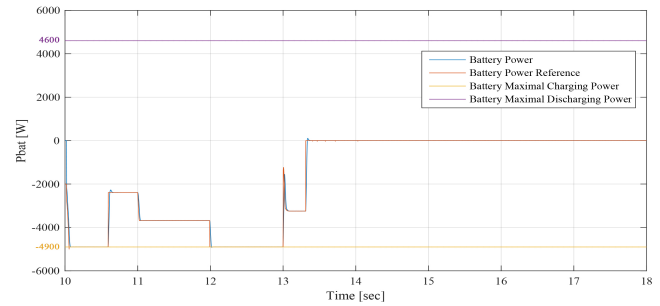


Figure 48: Battery and battery reference powers between 10sec and 18sec (EMS3)

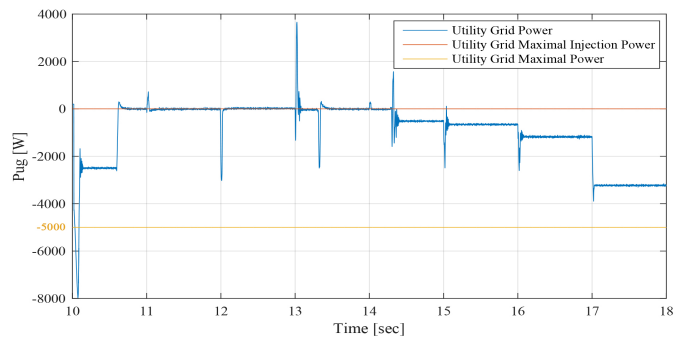


Figure 49: UG power between 10sec and 18sec (EMS 3)

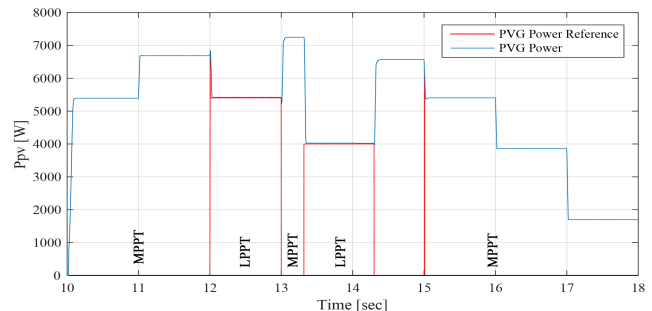


Figure 50: PVG and PVG reference powers between 10sec and 18sec (EMS 3)

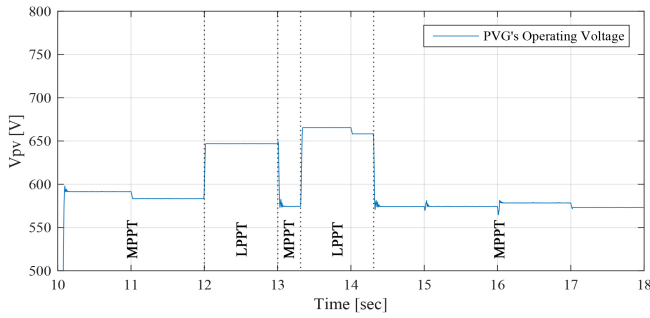


Figure 51: PVG operating voltage between 10sec and 18sec (EMS 3)

Opmode = 4 \rightarrow from $t=10s$ to $t=10.6s$: PVG generation increases with solar irradiation increase. It is sufficient to satisfy the load demand but still less than battery maximum charging power ($0 < P_{net} < |P_{bat}^{chmax}|$). Hence, the PVG operates under MPPT control. This surplus is used to charge the battery, since its SOC is lower than its maximum value SOC_{max} . And in order to ensure rapid charging of the battery (due to the SOC value inferior to 60%) aiming to obtain sufficient SOC for the next peak shaving request, the UG provides the necessary power, added to the surplus ($P_{UG} = P_{bat}^{chmax} + P_{net} < 0$), in order to charge the battery with its maximum power ($P_{batref} = P_{bat}^{chmax}$). Therefore, it is noticed that the SOC increases rapidly during this period.

Opmode = 3 \rightarrow from $t=10.6$ to $t=12s$ & from $t=13$ to $t=13.3s$: PVG is highly sufficient to satisfy the demand, but the surplus is still less than the battery maximum charging power ($0 < P_{net} < |P_{bat}^{chmax}|$). Hence, the PVG operates under MPPT control. This surplus is used to charge the battery, since its SOC is still lower than its maximum value SOC_{max} . The latter being quite high ($SOC > 60\%$), the UG power is not requested ($P_{UG} = 0$), and the battery is charged only with solar generation surplus ($P_{batref} = -P_{net}$), which does not exceed the maximum charging power of the battery. Thus, we notice that the SOC increases, and its evolution depends on the value of $-P_{net}$. Since SOC reached its maximum limit SOC_{max} , the battery charging is no longer authorized ($P_{batref} = 0$ at $t = 13.3s$).

Opmode = 2 \rightarrow from $t=12s$ to $t=13s$: PVG is largely sufficient to satisfy the demand, and the surplus is greater than the maximum charging power of the battery ($P_{net} > |P_{bat}^{chmax}|$). The SOC being less than its maximum value SOC_{max} , the battery starts charging with its maximum charging power ($P_{batref} = P_{bat}^{chmax}$). Since injection into UG is not authorized ($P_{UG} = 0$), The LimPV output takes the value 1 in order to switch to LPPT control. The PVG output power is limited to the value of $P_{pvref} = P_{load} + |P_{bat}^{chmax}| = 0.5kW + |-4.9kW| = 5.4kW$, calculated by the supervisor. The SOC increases rapidly as the battery charges with its maximum charging power. The voltage imposed by the LPPT algorithm is the one higher than the MPP voltage.

Opmode = 1 \rightarrow from $t=13.3s$ to $t=14.3s$: PVG is satisfying the demand ($P_{net} > 0$) and the battery SOC has reached its maximum value SOC_{max} , thus the battery can no longer be charged ($P_{batref} = 0$). As the injection into the UG is not authorized ($P_{UG} = 0$), the PVG is controlled by LPPT. The LimPV output takes the value 1 to switch toward the LPPT control. The PVG

reference power set by the supervisor is equal to $P_{pvref} = P_{load} = 4kW$. The battery remains at rest and its SOC equal to its maximum recommended (SOC_{max}), and the voltage imposed by the LPPT algorithm is the one higher than the MPP voltage.

Opmode = 7 \rightarrow from $t=14.3s$ to $t=18s$: Demand exceeds PVG production ($P_{net} < 0$), therefore the PVG is controlled by MPPT control. The deficit being less than the UG maximum power ($|P_{net}| < |P_{UG}^{max}|$), only the latter is requested to supports the PVG to meet demand ($P_{UG} = P_{net}$). The battery whose SOC is equal to its maximum SOC_{max} , remains at rest ($P_{batref} = 0$).

4.4. Overall System Results

Figure 52 shows that the overall system real power is equal to zero, which proves the performances of the GCHRES in terms of stability and power quality. Line to line voltages of the load is altered if the system contains real power [24]. From Figure 53, it is seen that the DC-bus voltage is inside standard limits [25], and Figure 54 shows the zero reactive power exchange between the ASC and the CCP. It should be noticed that the transients present in some figures as the ones corresponding to the UG power and the overall system net power, are due to the loads switching via circuit breakers in MATLAB/SIMULINK, and last no longer than tenth of second, which is negligible in the case of the real operation of the system throughout the 24 hours of the day.

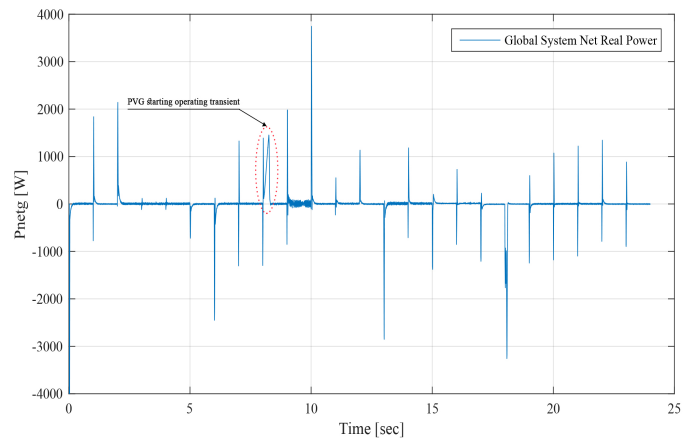


Figure 52: Overall system real power (simulation for EMS 1)

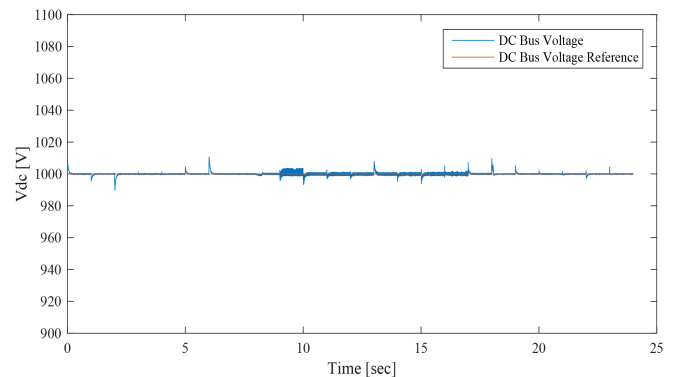


Figure 53: DC-Bus voltage vs reference (simulation for EMS 1)

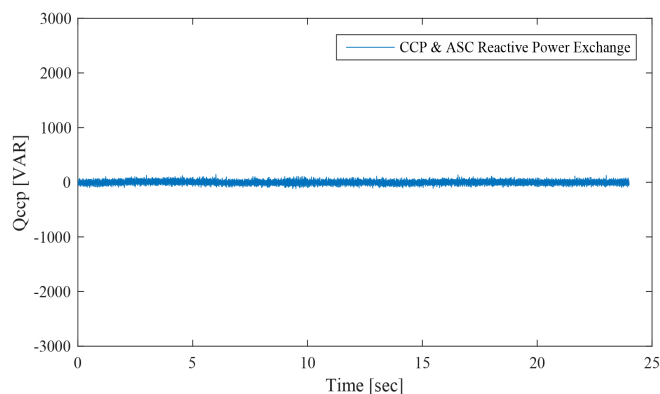


Figure 54: Reactive power exchange between ASC and CCP (simulation for EMS1)

5. Conclusion

This paper presents three novel energy management strategies controlling the power flows within a GCHRES. The proposed EMSs differ according to the UG metering types. They have for common purpose to continuously meet the variable demand of the load for the whole 24 Hours of the day. They aim also to avoid the battery lifespan reducing, and to reduce the monthly energy bill of the UG subscriber through battery peak shaving application. The dynamic behaviours of the proposed GCHRES, under the supervision of these several EMSs, are tested under real weather data and variable load demand profiles. The effectiveness of the developed system in terms of demand meeting, DC-Bus voltage regulation, overall system stability, power quality, and element constraints respecting, is confirmed by simulation in MATLAB/SIMULINK.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] K. H. Hussein, I. Muta, T. Hshino and M. Osakada, "Maximum photovoltaic power tracking: an algorithm for rapidly changing atmospheric conditions," *IEE Proceedings -Generation, Transmission and Distribution*, **142**(1), 59–64, January 1995, doi:10.1049/ip-gtd:19951577.
- [2] T. Wu, C. Chang, and Y. Chen, "A fuzzy - logic controlled single-stage converter for PV-powered lighting system application," *IEEE Transactions on Industrial Electronics*, **47**(2), 287–296, April 2000, doi:10.1109/41.836344.
- [3] C. Zhang, D. Zhao, J. Wang, and G. Chen, "A modified MPPT method with variable perturbation step for photovoltaic system," in *2009 IEEE 6th International Power Electronics and Motion Control Conference*, 2096–2099, 2009, doi:10.1109/ipemc.2009.5157744.
- [4] R. Faranda, S. Leva, V. Mageri, "MPPT techniques for PV systems: energetic and cost comparison," in *2008 IEEE Power and Energy Society General Meeting – Conversion and Delivery of Electrical Energy in the 21st Century*, 2008, doi:10.1109/pes.2008.4596156.
- [5] Patcharaprakiti N., Premrudeepreechacharn S., Sriuthaisiriwong Y, "Maximum power point tracking using adaptive fuzzy logic control for grid connected photovoltaic system," *Renewable Energy*, **30**(11), 1771–1788, 2005, doi:10.1016/j.renene.2004.11.018.
- [6] M. Dharif, Modélisation de l'intégration de l'énergie photovoltaïque au réseau électrique national, Ph. D Thesis, Cadi Ayyad University -Sciences and Technology Faculty of Marrakech, 2016.
- [7] S.-I. Go, S.-J. Ahn, J.-H. Choi, W.-W. Jung, S.-Y. Yun and I.-K. Song, "Simulation and Analysis of Existing MPPT Control Methods in a PV Generation System," *Journal of International Council on Electrical Engineering*, **1**(4), 446–451, 2011, doi:10.5370/jicee.2011.1.4.446.
- [8] N. Femia, G. Petrone, G. Spagnuolo and M. Vitelli, "Optimizing sampling

- rate of P&O MPPT technique," in *2004 IEEE 35th Annual Power Electronics Specialists Conference*, **3**, 1945-1949, June 2004, doi:10.1109/pesc.2004.1355415.
- [9] X. Liu, L. A. C. Lopez, "An improved perturbation and observation maximum power point tracking algorithm for PV arrays," in *2004 IEEE 35th Annual Power Electronics Specialists Conference*, **3**, 2005-2010, June 2004, doi:10.1109/pesc.2004.1355425.
- [10] A. Bennouna, "Autoproduction d'électricité-Pour qui est fait le Projet de Loi en gestation?," *chantiers du maroc*, **2** December 2020, <https://chantiersdumaroc.ma/btp-qui-bouge/dossiers/autoproduction-delectricite-renouvelable-la-loi-en-gestation/>.
- [11] European Parliament, "Outlook of Energy Storage Technologies", European Parliament's committee on Industry, Research and Energy (ITRE), 2008, IP/A/ITRE/FWC/2006-087/Lot4/C1/SC2. http://www.storiesproject.eu/docs/study_energy_storage_final.pdf.
- [12] Y. Riffonneau, Gestion des flux énergétiques dans un système photovoltaïque avec stockage connecté au réseau, Ph. D Thesis, Grenoble Alpes University, 2009.
- [13] A. Bennouna, "Produire son électricité solaire au Maroc, c'est possible et peut être rentable... malgré tout!," *EcoActu*, **30** March 2020, <https://www.ecoactu.ma/produire-son-electricite-solaire/>.
- [14] A. Guichi, A. Talha, EM. Berkouk, S. Mekhilef, "Energy Management and Performance Evaluation of Grid Connected PV-Battery Hybrid System with Inherent Control Scheme," *Sustainable Cities and Society*, **41**, 490–504, 2018, doi:10.1016/j.scs.2018.05.026.
- [15] P.B.S. Kiran, Limited Power Control of a Grid Connected Photovoltaic System, Ph. D Thesis, Indian Institute of Technology Gandhinagar, 2015.
- [16] A. Bouharchouch, EM. Berkouk, T. Ghennam, "Control and Energy Management of a Grid Connected Hybrid Energy System PV-Wind with Battery Energy Storage for Residential Applications," in *2013 Eighth International Conference and Exhibition on Ecological Vehicles and Renewable Energies (EVER)*, 2013, doi:10.1109/ever.2013.6521525.
- [17] J. Labbé, L'hydrogène électrolytique comme moyen de stockage d'électricité pour systèmes photovoltaïques isolés, Ph. D Thesis, Paris Mining School, 2006.
- [18] A. Dekkiche, Modèle de batterie générique et estimation de l'état de charge, Master in Electronic, Superior Technology school, Quebec University, 2008.
- [19] O. Tremblay, L. A. Dessaint, A. I. Dekkiche, "A Generic Battery Model for the Dynamic Simulation of Hybrid Electric Vehicles," in *2007 IEEE Vehicle Power and Propulsion Conference*, 284-289, 2007, doi:10.1109/vppc.2007.4544139.
- [20] R. A. Jackey, "A simple, Effective Lead-Acid Battery Modelling Process for Electrical System Component Selection," *SAE International*, **116**(7), 219–227, 2007, doi: <https://doi.org/10.4271/2007-01-0778>.
- [21] M. Ceraolo, "New dynamical models of Lead-Acid batteries," *IEEE Transactions on Powers Systems*, **15**(4), 1184–1190, 2000, doi:10.1109/59.898088.
- [22] SimPowerSystems TM Reference, Hydro-Québec/the Math Workds, Inc., Natick, MA, 2010.
- [23] European Commission website "ec.europa.eu", European Commission > EU Sciences Hub > PVGIS > Outils Interactifs, latest update 15 October 2019, https://re.jrc.ec.europa.eu/pvg_tools/fr/#MR
- [24] K. Tariq, H.S. Zulqadar, "Energy Management simulation of Photovoltaic/Hydrogen/Battery Hybrid Power System," *Advances in Science, Technology and Engineering Systems Journal*, **1**(2), 11–18, 2016, doi:10.25046/aj010203.
- [25] "IEEE Standard for Interconnecting Distributed Resources with Electric Power Systems," *IEEE Std 1547-2003*, 2003.

Appendix A

Figure A.1(a) shows real average solar irradiation and Figure A.1(b) the average temperature datas of the region of Marrakech, collected on hourly basis. Table A.1 shows the site localization and the data base informations.

Table A.1: Real solar irradiation and temperature data

Data base	PVGIS-SARAH
Region	Marrakech / Morocco
Latitude / Longitude	31,627 / - 7,988
Month	January
Horizon	calculated
Irradiation	Fixed orientation /Tilt 30°/Azimuth 0°

Appendix B

Table B.1 shows the maximum value of charging/discharging power that the battery is supposed to receive/provide, for each month, calculated respectively by (1) and (2).

$$\max_{t=0...24} (P_{PV}^{MPPT}(t) - P_{load}(t) + P_{UG}^{max}) \quad (kW) \quad (1)$$

$$\min_{t=0...24} (P_{PV}^{MPPT}(t) - P_{load}(t) + P_{UG}^{max}) \quad (kW) \quad (2)$$

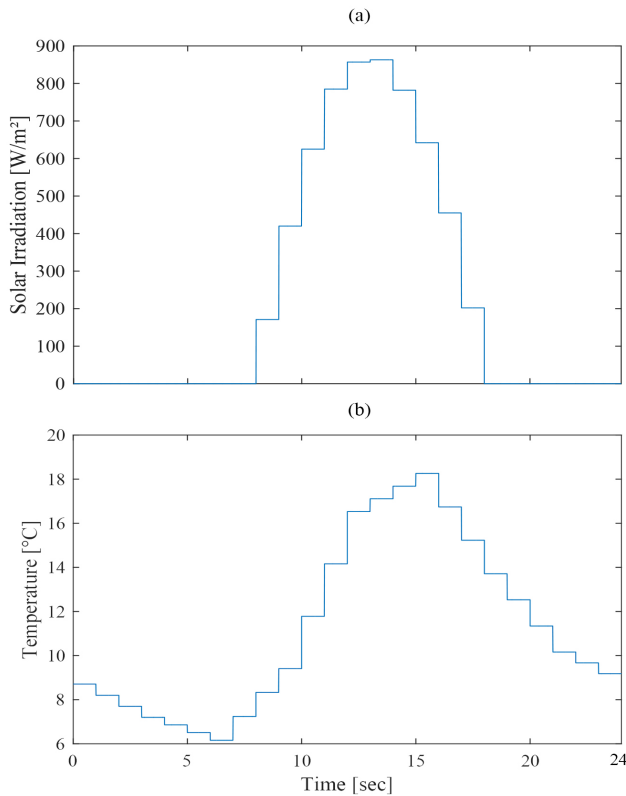


Figure A.1: Real Weather data (a) average solar irradiation (b) average temperature (January-Marrakech)

absorb/provide during a day, for each month, calculated respectively by (3) and (4).

$$\sum_{t=1}^{24} \sigma (P_{PV}^{MPPT}(t) - P_{load}(t) + P_{UG}^{max}) \quad (kWh) \quad (3)$$

$$\sum_{t=1}^{24} \lambda (P_{PV}^{MPPT}(t) - P_{load}(t) + P_{UG}^{max}) \quad (kWh) \quad (4)$$

Table B.1: maximum charging and discharging power of the battery of each month

Month	Max Charging Power	Max Discharging Power
Jan	4.204	-4.600
Feb	4.436	-4.564
Mar	4.694	-4.432
Apr	4.886	-4.348
May	4.534	-4.234
Jun	4.662	-4.132
Jul	4.830	-4.006
Aug	4.874	-4.000
Sep	4.640	-4.162
Oct	4.272	-4.342
Nov	3.654	-4.450
Dec	3.931	-4.558

Table B.2: Maximum absorbing and providing energy of the battery of each month

Month	Max Absorbing Energy	Max Providing Energy
Jan	19.328	-29.649
Feb	20.946	-27.747
Mar	24.068	-22.700
Apr	26.322	-18.922
May	25.093	-16.231
Jun	27.539	-15.463
Jul	29.800	-14.860
Aug	29.983	-16.291
Sep	25.282	-18.835
Oct	20.332	-21.244
Nov	16.031	-26.319
Dec	17.350	-28.481

And Table B.2 shows the maximum value of charging/discharging energy that the battery must be able to

IT Project Management Models in an Era of Digital Transformation: A Study by Practice

Rachida Hassani*, Younès El Bouzekri El Idrissi

Systems Engineering Laboratory, National School of Applied Sciences – IBNT TOFAIL University, Kénitra, Morocco

ARTICLE INFO

Article history:

Received: 23 December, 2021

Accepted: 24 February, 2022

Online: 18 March, 2022

Keywords:

IT project

IT project management

Waterfall

Agile Hybrid

Digital transformation IT project models

ABSTRACT

In 1998 we were talking about an NTIC era, and since the years 2010, we are tracing the beginning of a new era of technological development, called the era of digital transformation. During this new era companies have started to run towards the digitalization of their processes, which represent a large part of the market, the implementation of IT projects for the simplification and automation of their business. This new technological mode has generated new constraints related to the digital transformation in the management of IT projects. Beyond the constraints related to this era, old constraints already exist in the literature in the two modes of development that can be found, waterfall or agile. The main purpose of this study is to review the literature for a theoretical understanding of the issue, and an analysis through practice, through a case study and observations in the field, to better identify and understand the practical concerns. The results of this study led us to identify the constraints of digital transformation, the strong and weak points of each management model. The matrix linking these elements allowed us to propose a new hybrid development mode that exploits the potential of both waterfall and agile models, while considering and highlighting the constraints of this new era.

1. Introduction

Over the last four decades, numerous IT development models and project management standards have emerged. However, with the evolution of technology, companies specializing in IT development have encountered serious problems, which have gradually pushed them to reconsider their adopted methods of development and IT project management. This reconsideration becomes more and more urgent with the exponential evolution of technologies, from the era of new information and communication technologies "NICT" to the era of digital transformation "DT".

In the late 1990s, IT project management methods based on a waterfall lifecycle approach were the most dominant methods [1]. The most representative approach to this type of development is the "cascade" model. However, over the last few years, the principles of linear breakdown and IT project management based on "Waterfall" methods have been increasingly called into question. The permanent change in the functionalities to be developed and the "tunnel" effect caused by "cascade" projects have led practitioners to question this development model and to focus on the rapid and constant evolution of the technology market.

In very few years, the concept of agility has become a major success in the IT development industry. According to a survey conducted by the PMI in 2017, 71% of the organizations surveyed said they use agile approaches for their projects sometimes, often or always [2].

Despite the emergence of this new mode of development, which makes it possible to cover the shortcomings of waterfall models, and despite its exponentially increasing rate of use, the failure rate of IT projects remains one of the main concerns of companies. Indeed, according to the latest results of the Standish Group Chaos study published in their last report in 2018, only 34% of projects are successful. While 51.5% are challenged and 14.5% are cancelled before completion. That said, 66% of developed projects are partially or completely failed [3].

Despite the abundant literature on the concept of agility, the transition to an "agile" organization is a real challenge. At present, the context of the use of "agile" tools remains unclear. The analysis of the literature published on the subject underlines the contradictory positions of practitioners with regard to the applicability of the management practices and instruments carried by these approaches. Agile methods articulate new managerial concepts and devices without, however, participating in a unified management model. They do not seem to be based on a structured project management approach. They thus raise the question of the feasibility degree of the technical substrate associated with them, of the greater or lesser relevance of the management philosophy that they underlie, and of their internal contextualization [2].

While the application of these methods has received very positive feedback at the level of small teams, the results have been less telling in terms of their application in complex organizational structures, without mentioning the lack of project management

*Corresponding Author: Rachida Hassani, rachida.hassani22@gmail.com

skills within these teams and their ability to properly self-manage the content of each iteration. Among the empirical studies that have dealt with the implementation of these "managerial innovations", few have focused on their application in "complex" organizational structures, and especially in an era of digital transformation. As a result, we note a real gap in the literature devoted to this subject.

The philosophy of "agile" methods thus seems to pose in a central way the question of collective sense, and this in an interactionist perspective [4] where the sensemaking is done in a processual way through communication between the participants. From this point of view, the organizing dynamics to which these methods refer can be understood as continuous sequences of interactions between the project's actors. This leads us to wonder how this type of "managerial innovation" [2] can be implemented in a structured management approach to constitute a formalized organizing system.

The era of digital transformation is characterized by the need to produce IT tools very quickly to keep pace with competitors, and which requires a very high level of quality within very precise deadlines and in an exponentially changing environment that once again demands a very high level of adaptability and flexibility. To cope with these new requirements of this era, companies have been running towards agility that does not respond properly to the constraints of this new era, and they have resorted to a new mode of development that is not yet mature, and which they have called false agility. This new notion refers to the adoption of an agile project management model, while keeping the contractual aspects found in traditional management models. This need is linked on the one hand to the new requirements of this era of digital transformation, and on the other hand to the lack of mastery of agility in practice.

The purpose of this article is to respond to this research problem by articulating a critical analysis of the literature devoted to waterfall and agile methods and a case study conducted from the perspective of the "by doing" approach [5]-[9]. The latter focuses on human action to understand the functioning of groups and their links with the organization and society [10].

The objective of the observations made is to analyze the process of "manufacturing" a project management strategy adapted to this new era of digital transformation and its implementation in an organization characterized by unstable teams, attached to several projects and geographically dispersed.

2. Literature review

2.1. IT project and IT project management

2.1.1. IT project

A project is a set of coordinated activities and actions that make it possible to respond to a precise and clearly identified need, by mobilizing resources and respecting a deadline [11]. It is composed of a series of tasks that have one and the same objective. These tasks are subject to conditions, including the time, people and resources needed to complete them [12].

An IT project is a project whose deliverables consist of IT tools, methods, or services, it is characterized by its uniqueness and punctuality. Typically, IT projects consist of five steps:

www.astesj.com

initiation, planning, execution, monitoring and closure. Each step is composed of specific tasks that enable the achievement of project objectives.

2.2. IT project management

Project management is the art of managing one or more projects, and consists of providing the means of prevention, detection, and analysis to ensure, throughout the project, the best possible match between objectives, costs, and deadlines. Managing a project means knowing where you are going (objectives), how you are going to get there (budget and resources) and when you are going to get there (deadline). It is about building and maintaining a real information system around the project management.

To achieve this, we must have at our disposal means such as a methodology for breaking down the project into tasks, techniques for calculating deadlines, cost and budget plans, a methodology for periodically assessing progress, progress metrics, etc. It is all the operational and tactical aspects that make a project end up in a triangle representing the quality-cost-delivery balance (QCD). All these means implemented will enable the organization, forecasting, monitoring and analysis of the project's progress.

2.3. Project management challenges in the era of digital transformation

The digital transformation presents several challenges for businesses. More specifically in terms of project management. The technological environment has evolved, the era of transformation is growing. However, project management processes remain frozen, no adaptation of project management methods has taken place to circumvent the requirements of this new way of doing business. Indeed, the difficulty of the digital transformation is furthermore since large global organizations often have traditional technologies and well-established work methods. In addition, they also have third-party partners, which adds to their complexity and therefore makes them vulnerable to competition, especially from smaller, more agile, and digitally focused start-ups.

Beyond the constraints listed above, digital transformation adds a new layer of complexity, linked to the fast-growing technological environment, with very specific IT processing and production deadlines, which require complete upstream project planning, with a well-defined scope and main objectives, but which may be subject to change.

2.4. IT project management methods

2.4.1. Waterfall methods

The waterfall model of V-Cycle (or Cascade) is a sequential and linear management method that allows to represent the developments through successive phases (figure 1) [13]. It is divided into several steps identified from the start of the project. The validation of one step leads to the start of the next, without any overlap of steps. This method is mainly used in software development, it limits the returns to previous stages, requires the business to define a long-term objective, and focuses on end-to-end project planning and is resistant to change, they are called "predictive".

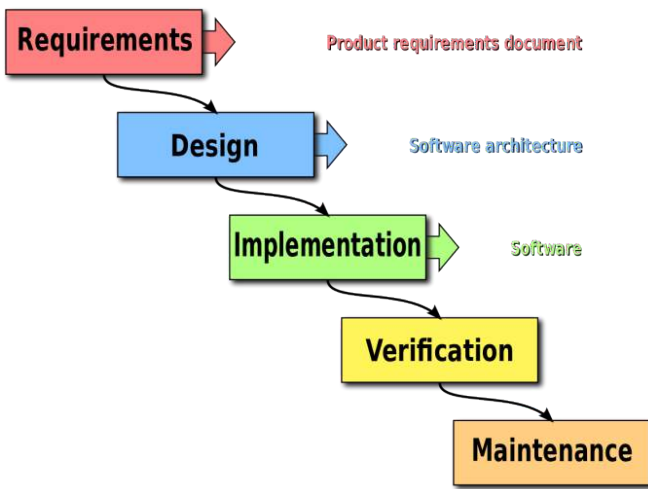


Figure 1: Project Management Methods: Waterfall Method

These methods are characterized, on the one hand, by their sequential aspect and, on the other hand, by the "customer/supplier" relationship between the project owner and the project manager. They induce a real rigidity in the project management.

By applying this methodology, the project manager must then commit to a precise schedule for the project's realization by foreseeing milestones for the beginning and end of steps as well as the tasks to be carried out. The project team commits to this precise schedule and defines all the tasks to be carried out. Each task is implemented on the following one, with a mandatory dependency, to reach the expected deliverable.

This methodology requires very precise specifications, tasks that are meticulously anticipated and described, with little chance for the unexpected. The project team follows these specifications to the letter and works on the entire project until its delivery. The project is therefore managed in its entirety, with little customer or external interaction during the production step.

2.4.2. Agile methods

Agile methods proceed in stages with short-term objectives (figure 2). They use an iterative development principle which consists in dividing the project into several stages called "iterations" (14). These iterations are nothing more than mini projects defined with the client detailing the different functionalities that will be developed according to their priority. "The project manager then establishes a macro-planning corresponding to the tasks necessary for the development of these functionalities". The goal is to assume that we cannot know and anticipate everything no matter how much experience we have. We then cut the project into iterations rather than anticipate and plan for everything, knowing that unforeseen events will occur along the way.

The Agile methodology is generally opposed to the waterfall methodology. It is intended to be more flexible and adapted and places the customer's needs at the center of the project's priorities.

In the field of IT development, a certain number of practices have been established, allowing the implementation of agility in a team or a company, the notion of "agile development" remaining very general. Many of these methods can be found in variants of

agile software development, such as Scrum, Kanban, extreme programming (XP), Feature-driven development, Behaviour Driven Development or Crystal:

- Backlog
- Retrospective
- User story
- Agile testing
- Programming in pairs
- Time Box

There are many other methods in agile methods. They have in common that they aim at improving the efficiency and quality of work.

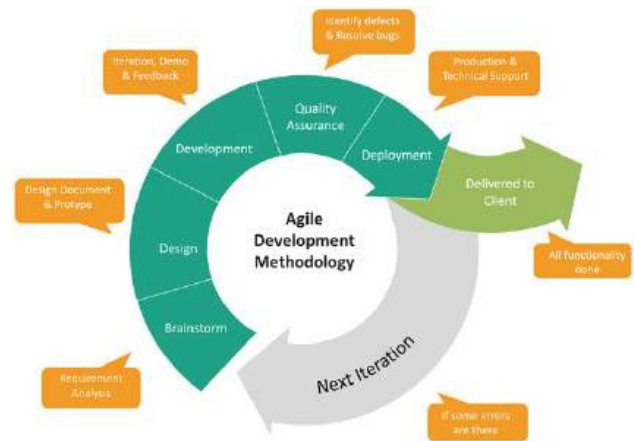


Figure 2: Project management methods: Agile method

An "iterative" development prioritizes user requirements that may evolve during the project, while developing a relationship between product owner and sponsor of working partners with common objectives.

The agile methodology is suitable for projects that are not too strict on deliverables and above all that require real-time adaptation to customer needs and requirements. It is becoming more and more common, but requires good working habits, especially in the team members communication, in real time. This methodology allows the improvement of:

- The quality of communication.
- Visibility.
- Quality control.
- Risk detection.
- Motivation and confidence of the team.
- Cost control.

2.4.3. Analysis and comparative study of waterfall and agile management methods

By closely studying and analyzing both classical and agile methods, we can summarize the differences and characteristics of the two methods in the following table (table 1):

Table 1: Analysis and comparative study of classical vs. agile methods

Topic	Waterfall	Agile
Risk Management	Separate, rigorous risk management process.	Risk management integrated into the overall process, with accountability for identifying and resolving risks. Risk-based management.
Success Measuring	Respect of initial commitments in terms of costs, budget and quality level.	Customer satisfaction through the delivery of added value and speed of implementation.
Lifecycle	Cascading or V- shaped Key sequential steps. This model makes parameter rollbacks very tricky.	Iterative and incremental Allows adjustments Accepts changes during iterations Each iteration delivers a usable part of the solution.
Planning	Predictive Detailed plans based on a perimeter and requirements defined and stable at the beginning of the project.	Adaptive with several levels of planning (macro and micro-planning) with the possibility of adjustments if necessary, as changes occur.
Documentation	Produced in large quantities at each stage as a support for communication, validation and congratulation.	Reduced to the main essentials in favor of functional, operational increments to obtain customer feedback.
Team	A team with specialized resources, led by a project manager.	An empowered team where initiative, versatility and communication are privileged, supported by the project manager.
Quality	Quality control at the end of the development cycle, once the product is finished. The customer discovers the finished product at the end of the project.	Early quality control, regular at each iteration, and permanent, at product and process level. The customer visualizes and validates the results early, on a continuous basis, and remains involved in the control frequently.
Change	Resistance or even opposition to change. Cumbersome process of managing accepted changes.	Favourable reception of the inevitable change, integrated into the process.
Follow-up on progress	Measurement of compliance with initial plans: quality, cost and lead time. Gap analysis.	Only one progress indicator: the number of functionalities implemented and the rest to be done.

2.5. Epistemological choices and research approach

2.5.1. Epistemological choices

This article aims to dissect the techniques and practices of IT project management within companies specialised in IT production (the case of digital service companies). To do so, we first opted for an ethnographic study, i.e. a participant observation of a human process, a set of tasks and events in a particular site through a longitudinal data collection. Adherence to such a perspective already pushes us towards an exploratory type of research and to opt for a qualitative approach. The researcher then wants to continuously and systematically induce a meaning of the events he/she observes.

The study of IT project management models lends itself well to the ethnographic perspective. This is referred to as an idiographic rather than nomothetic research strategy.

2.5.2. The research approach

Our research approach aims to identify traditional (Waterfall) and agile IT project management models, in an era of digital transformation, for the case of outsourcing (the case of digital

services companies). We used the inductive method because although we had already thought about what we thought was a likely situation, we did not have a well-articulated theory a priori. We therefore limited ourselves to observing a situation, understanding its meaning and trying, from this new understanding, to induce, if possible, a certain theory.

This choice of the case study technique was made with full knowledge of the facts, since we know that it offers great advantages such as the in-depth analysis of a site, the possibility of developing historical parameters, and a strong internal validity. It is ultimately an adaptable technique.

2.6. The steps of the research

We chose the multiple case study as our research technique. The value of such a technique depends greatly on the researcher's ability to demonstrate the credibility of his or her approach. In order to put all the chances on our side, we have respected and followed, very scrupulously, the different steps generally adopted to respond to this research technique:

3. Methodology

3.1. Data collection

Data collection for each of the cases studied was done mainly through participant observation.

3.1.1. Data processing

In multiple case studies, it is good practice to develop a schema that will be used to code the data. This scheme (figure 3) is also a means of ensuring a consistent method of comparing data from different case studies.

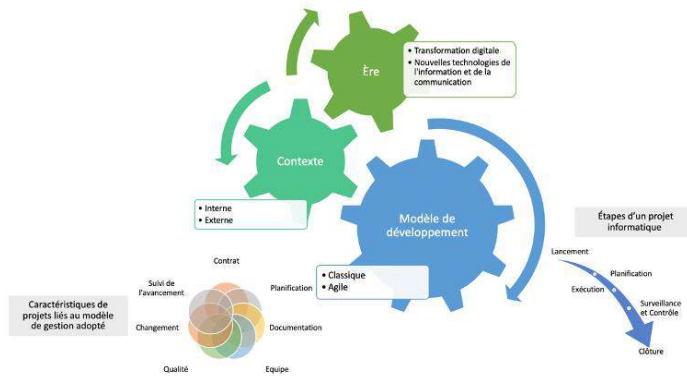


Figure 3: Schéma de codification utilisé pour le traitement des données.

3.1.2. Interpretation of the data

For the interpretation of the data, we linked the results of the coding of the raw data of each case. This is how we were able to see, for each dimension that makes up the proposed thematic grid, whether there were certain trends that emerged in the three codified cases. From this first globalisation, we were able to highlight the convergences between the traditional and agile management models in this era of digital transformation in all the cases, i.e. what works and what does not work in this era of digital transformation in each of the models used.

3.1.3. The study population and the sample selected

One of the important steps in the multiple case study is to determine the area that is affected by the research. This means identifying, from clearly defined variables, the population under study. It is also necessary to state the precise rules from which we have chosen, from this population, a sample of cases that has been the object of our observation.

Three major variables are part of our problematic: the project management model (agile and classical), the era in which projects are produced (digital transformation era) and the production context (external).

3.2. Methodological motivations

Our research project aims to understand the modes of "construction" of a classical and an agile approach, while taking into consideration the new characteristics of IT projects, specifically in this era of digital transformation. The "practice" perspective can be considered as a relevant research angle to examine, in depth, the activities of the actors involved in the "making" of these project management methods and their continuous interactions with the environment in which they are

located, as well as the tools and practices used in each of the phases of the life cycle of an IT project in general.

To date, no study has been conducted on how the actors collectively interpret and construct methods adapted to the constraints of this new era, which must be implemented in "complex" organizational structures. In the two "toolboxes" of classical and agile methods, which instrumentation(s) should be favoured? How do the elements of organizational contingency intervene in the implementation of the engineering and managerial principles specific to this era?

To answer these questions, it seems appropriate to us to focus on a field study to better understand how tools classified under the term "era of digital transformation" are implemented within organizations. The choice of an instrumental case study thus seems relevant to us to analyze, in depth, the phenomenon of implementing methods that consider the new constraints of this era in "complex" organizational structures.

The case method is a precise mode of observation of themes previously defined by questioning (15). It supposes an in-depth analysis of the various aspects of a situation to reveal the significant elements and the links between them. Moreover, case studies can be classified in three categories (16): intrinsic, instrumental, and multiple case studies.

Within the framework of this work, we wish to carry out a case study of instrumental type in organizations specialized in the production of IT projects in an era of digital transformation.

This approach gives us on the one hand a deep understanding of IT project management, specifically in this new era of digital transformation, and on the other hand the identification of recurring phenomena during the production of IT projects, to identify the requirements of this new era. It also consists in identifying the most recurring phenomena when using classical and agile methods, to identify what works and what does not work in each method in this new era. Our intention is to capture the way classical and agile practices are implemented in this new era as project management methods. To do so, we have decided to approach the selected cases from a "by doing" perspective, focusing on what the actors in charge of implementing project management tools do.

3.3. Presentation of the cases of the studied companies

The 3 companies studied are companies specialized in the production of IT projects, specifically web and mobile projects (information systems, web software, websites, e-commerce sites, etc.).

The duration of the study of each company was limited to one year, which gives feedback of the experiences of the 3 companies for a global period of 3 years. Also, the choice to study 3 different companies allows us to generalize the studied phenomenon as well as the contributions made within the framework of this article.

3.3.1. Case A

This company adopts the project management method imposed by its clients (the sponsors). Projects that adopt a waterfall model do not undergo strong changes compared to the definitions in the literature.

The projects that adopt an agile model undergo changes according to the constraints of each client and which is generally summarized by:

- definition of a fixed project cost
- definition of a specific contractual aspect of the project
- imposition of a precise schedule according to the customer's constraints
- deleting or adapting the following events:
 - replace the daily meeting by a weekly team point
 - definition of the backlog by the project manager in the form of an excel file with a well-defined structure.
 - nonhomogeneous iterations in terms of time interval
 - a content of each iteration that is fixed, and that implies a shift of the next iteration if the delays cannot be maintained

For corrective maintenance projects, adoption of an iterative agile model removing the following events:

- Retrospective
- Demonstration

While including the customer constraints defined above.

The following table (table 2) summarizes the methods used, highlighting the size of the projects, the project failure rate, and the list of main causes:

Table 2: Failure rate statistics of case A

Method	Size	Number	Failure rate	Causes of fails
Waterfall	Small	29	10%	Team & Communication
	Medium & Large	67	30%	Cost Quality Change Team & Communication
Agile	Small	19	45%	Planning Documentation Progress tracking
	Medium & Large	5	100%	Planning Cost Documentation

3.3.2. Case B

This company imposes an agile management method on all its customers (the sponsors).

The projects that adopt an agile model undergo changes according to the constraints of each client and which is generally summarized by:

- definition of a fixed project cost
- definition of a specific contractual aspect of the project
- imposition of a partial precise schedule according to the customer's constraints
- deleting or adapting the following events:
 - deletion of the retrospective
 - addition of two events, project committee (weekly), and steering committee (monthly)
 - definition of the backlog by the project manager in the form of an excel file with a well-defined structure.

The following table (table 3) summarizes the methods used, highlighting the size of the projects, the project failure rate, and the list of main causes:

Table 3: Failure rate statistics of case B

Method	Size	Number	Failure rate	Causes
Agile	Small	3	33%	Planning Documentation
	Medium & Large	13	55%	Planning Documentation Cost Progress tracking

3.3.3. Case C

This company adopts the project management method imposed by its clients (the sponsors).

Projects that adopt a waterfall model do not undergo strong changes compared to the definitions in the literature.

The projects that adopt an agile model undergo changes according to the constraints of each client and which is generally summarized by:

- definition of a fixed project cost
- definition of a specific contractual aspect of the project
- imposition of a precise schedule according to the customer's constraints
- deleting or adapting the following events:
 - definition of the backlog by the project manager in the form of an excel file with a well-defined structure.
 - non homogeneous iterations in terms of time interval
 - a content of each iteration that is fixed, and that implies a shift of the next iteration if the delays cannot be maintained
 - addition of two events, project committee (weekly), and steering committee (monthly)

For corrective maintenance projects, adoption of an iterative agile model removing the following events:

- Retrospective
- Demonstration

While including the customer constraints defined above.

The following table (table 4) summarizes the methods used, highlighting the size of the projects, the project failure rate, and the list of main causes:

Table 4: Failure rate statistics of case C

Method	Size	Number	Failure rate	Causes
Waterfall	Small	7	42%	Team Progress tracking
	Medium & Large	13	70%	Change Team & Communication Quality
Agile	Small	--	--	--
	Medium & Large	6	50%	Planning Documentation

3.4. Investigation process

Our qualitative approach mobilizes a set of data collection and analysis techniques to identify, understand and interpret events in their context.

4. Results

The results of our investigation can be summarized as follows:

4.1. Case of projects following a waterfall model

Project size has a direct impact on the success rate of IT projects. Indeed, if the project is small, the success rate increases and with the increase of its size the risk of its failure increases. A small project is therefore a project that is simple to manage and control.

Communication in waterfall projects is very poor and has a very negative impact on the continuity and smooth running of the project. The rollback process in any project step is too risky or even impossible when it comes to advanced steps.

The detailed planning gives a precise vision on the continuation of the project, the customer / sponsor can afford to prepare activities that are related to the project by defining the deadlines in advance.

Project documentation always has an impact on the success of this type of project. However, in the same company, when there is no writing standard, there are very important quality gaps that depend on the project manager who wrote the documentation and his experience. Similarly, this increases the risk of oversights and major gaps in the requirements specification.

Quality control is only carried out at the end of the development process, which means that no quality risk management is implemented. The customer can only identify functionalities that do not or no longer meet his needs at the end of the project, with a very costly backtracking at this stage of the project.

This type of projects cannot accept any type of change, this is one of the main constraints of IT development in an era of digital transformation, technological evolution runs at the speed of light, and therefore customer needs can be subject to evolution and change and full process of project implementation. These changes are in opposition to this mode of development, they are expensive, even impossible in some cases, which pushes some customers to cancel their project before its completion, to start again another one that meets the new requirements, again taking the risk of throwing this project away if the needs evolve overtime.

Exclusion of the customer during the different steps of production, therefore no risk can be anticipated, which generates customer dissatisfaction, and even the production of projects that will not be used afterwards, because it does not reflect the vision of the project on the customer's side and this risk has not been identified in the primary steps. For projects managed according to a waterfall model, the presence of exhaustive and up-to-date project documentation has a direct impact on the simplification of its maintainability, handling, and subsequent scalability.

4.2. Case of projects following an agile model

The planning methods adopted in practice are not mastered or even understood by team members. In the same way, the macro planning of this development mode is never fixed in terms of time and is not respected either, this is linked to the fact that important functionalities are shifted from one iteration to another, especially in the first iterations where the production capacity is not yet mature. These shifts are indirectly responsible for wastage linked to the planning of tasks and activities that are external to development, but which depend on them at the same time. That said, if we look at the project only it can be considered as successful, but if we count the damages linked to its bad planning, the results are always catastrophic.

The documentation is always light, ideas are mostly developed in real time between the development team, the product owner, and the customer himself, through questions and requests for clarification. On the other hand, the latter are never documented, so the project evolves in the right direction, but its documentation does not follow this evolution and remains very limited.

Agile teams always suffer from project backlog that are not complete or very poor in terms of detail, which complicates their management of time, understanding, as well as production.

This type of documentation is considered in practice as one of the major problems that lead to the failure of IT projects, in terms of scope, quality, cost and planning. It has also proved its limitations in the maintenance steps, where new teams cannot find a written record of what has been produced, especially for projects where specific management rules are defined, and indeed it is most IT projects that require specific management rules. This complicates the maintainability of agile projects, or even makes it impossible to upgrade and stabilize them.

Agile project teams are self-managing, but in practice, developers and technical teams do not have the project management skills, making risk management and anticipation virtually nil. Decision-making during production steps does not necessarily reflect a relevant strategic and project management logic.

Progress indicators in agile management focus on the number of functionalities developed and what remains to be done. In practice, the contractual aspects required by customers add another layer that concerns the commitment to results, deadlines and quality. And therefore, a return to the 3 main constraints of IT development quality, cost, planning, by adding new constraints related to the scope, added value, and scalability of projects as well as other constraints related to this new era of digital transformation and the specific needs of each customer. That said, lack of freedom

over the other variables that become fixed and binding constants for the company responsible for developing the IT solution.

5. Discussions

The analysis of these research results has allowed us to identify what works in practice in the different waterfall and agile development modes. It has also allowed us to put our finger on the constraints of this new era called the era of digital transformation, to propose a new development mode that highlights the 3 elements studied (Waterfall mode, Agile mode, Era of digital transformation).

Before proposing a new development mode, it is essential to define a matrix that highlights what works in each of the studied development modes (table 5):

Table 5: Constraint matrix of waterfall model, agile model, and the era of digital transformation

Topic	Waterfall	Agile	Things to keep in mind
Lifecycle	Perfect life cycle for small projects	Iterative life cycle, allows to divide the project into small projects.	Divide each project into small projects equal in terms of the effort required for each iteration. => How to do this division?
Planning	Definition of a detailed schedule	Feedback to identify the realistic production capacity of each iteration.	For each iteration produces a detailed planning (classic mode). Consider the feedback of the previous iteration to fix the perimeter of the next iteration. Each iteration must include a list of prioritized features according to the MoSCoW principle. This principle gives a risk management variable on the functionalities that can wait during each iteration without impacting the macro planning.
Documentation	Important for each step and allows to define the communication, validation, and contractual aspects	Production of a validated backlog before the start of each iteration	Production of a detailed backlog of each iteration (documentation of the waterfall mode). The backlog of each iteration is fixed, only the project manager and the customer can modify it. Modifications must be of type: cancellation of functionalities subject to evolution and change (to be put on standby in the current iteration) Identification of the most urgent features of the next iteration that are already validated
Team	A team with specialized resources, led by a project manager.	An empowered team where initiative, versatility and communication are privileged, supported by the project manager.	At the head of the team a project leader, (conductor) to manage the project and not the team. The project leader will enable the implementation of project management strategies, priorities, and alignment of the teams on the project roadmap. He will also have the role of simplification for the development teams. The teams are also the technical managers of the project, and must in their view alert, propose, communicate, and warn if necessary. Each team member, including the project manager, is responsible for the entire iteration and not for a part of the

			iteration or a particular task within the iteration.
Quality	--	Writing of the test document in parallel with the writing of the backlog.	Production of the quality definition document in parallel with the specifications of each step.
		Each iteration ends with a validation demonstration with the project owner and the customer.	Each functionality requires a continuous control within the team with a first validation by the project manager. Each iteration requires complete control and validation by the team, the project manager and the customer.
Change	--	Favorable reception	Favorable reception but following processes, so as not to endanger the current iteration.
			Each change must be studied by the client / project manager.
			Once the decision is made between this pair, it must be validated with the development teams.
			Once validated, this requires an arbitration to cancel the functionality subject to the change in the current iteration (if the change is heavy), and replace it by an urgent, detailed, and validated functionality of the next iteration (if the production capacity of the current iteration allows it).
Follow-up on progress	Measurement of compliance with initial plans: quality, cost and lead time. Gap analysis.	Indicator of what remains to be done	Continuous analysis of what remains to be done in relation to the scope of each iteration, which allows anticipation of risks related to contractual aspects and quality, cost and deadline compliance, with continuous gap analysis that will guide managers in making project decisions.
Risk management	--	Risk management integrated into the overall process, with accountability for identifying and resolving risks. Risk-based management.	The project evolves, and the risks evolve in parallel. This implies a risk management plan integrated into the overall process, with everyone being accountable for identifying and resolving risks for each iteration.
Measuring Success	Respect of initial commitment in terms of costs, budget and quality level.	Customer satisfaction through the delivery of added value and speed of implementation.	Compliance with initial contractual commitments. Production of added value with continuous integration. Customer satisfaction.

To respond to the constraints of this new era of digital transformation, and to face the limits of mastery of development modes observed mainly in agile models, as well as the contractual aspects not dealt with in agile modes, but which are required for computer production companies. It is essential to propose a hybrid development mode, which highlights the strengths of each mode (what works), as well as the constraints of this new era (requirements linked to continuous and rapid technological change). This hybrid mode (figure 4) can be defined as follows:

6. Conclusion

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the Microsoft Word, Letter file.

In this article we deal with the issue of IT project management in an era of digital transformation. To do so, a study of the literature was essential to understand project management as a managerial innovation in the literature, and as a complement to this study, a practical application in the field which lasted 3 years in 3 IT

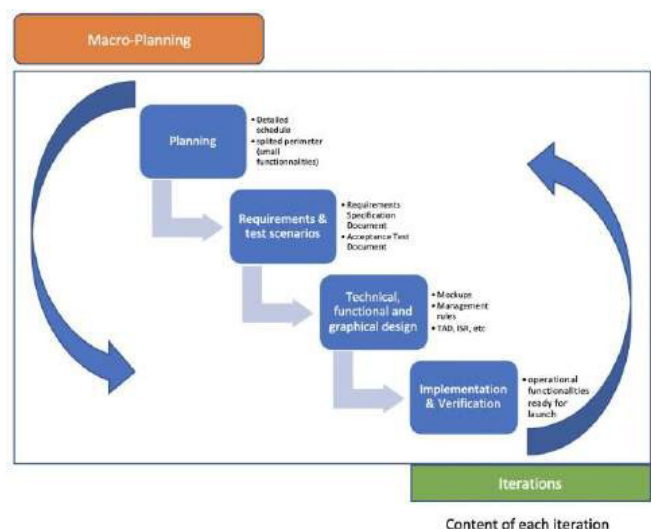


Figure 4: Hybrid mode for IT project management in the era of digital transformation

In this study, we put our finger on what works and what doesn't work in waterfall and agile project management methods. We have production companies. The objective of this study through practice is to understand the environment, links and interactions between the different entities that make up the IT project management ecosystem added to this matrix the constraints related to the digital transformation era.

The result of matching between different entries of this matrix, allowed us to define the outline of a new development mode, which we called hybrid, because it is composed of a mix between waterfall and agile practices.

7. Conflicts of Interest

The authors confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

8. Acknowledgment

National School of Applied Sciences of Kenitra, IBN TOFAIL University, Morocco

9. References

- [1] T. Nonaka, H. Nonaka, I. Nonaka, "The new product development game", *Harvard Business Review*, 137-146, 1986.
- [2] P. M. Institute, "Success Rates Rise - Transforming the high cost of low performance", *Global Project Management Survey*, 2017.
- [3] J. Johnson, "CHAOS Report: Decision Latency Theory - It Is All About the Interval", *Standish Group International*, 2018.
- [4] A. David, "Structure et dynamique des innovations managériales", *Ecole des mines*, 1996.
- [5] W. K. E. V. Bénédicte, "Le sens de l'action: Karl E. Weick, sociopsychologie de l'organisation", Paris: Institut Vital Roux, 2003.
- [6] H. Mintzberg, "Le manager au quotidien - Les 10 rôles du cadre", *Organisation editions*, 2002.
- [7] H. A. R. L. D. Albert, "Les nouvelles fondations des sciences de gestion: Éléments d'épistémologie de la recherche en management", *Presse des Mines*, 1er édition, 2012.
- [8] S. Gherardi, "Practice-Based Theorizing on Learning and Knowing in Organizations", *Organization*, **1**(27), 211-223, 2000, doi:10.1177/135050840072001.
- [9] R. Whittington, "The Work of Strategizing and Organizing: For a Practice Perspective", *Strategic Organization*, **1**(21), 117-125, 2003, doi: 10.1177/147612700311006.
- [10] R. Whittington, "Completing the Practice Turn in Strategy Research", *Organization Studies*, **1**(5), 613-634, 2006, doi: 10.1177/0170840606064101.
- [11] E. Antonacopoulou, "Strategizing as Practising: Strategic Learning as a Source of Connection", *Advanced Institute of Management Research (AIM) Research Paper Series*, 1-35, 2006, doi: 10.2139/ssrn.1307066.
- [12] F. A.-P. V. W. L. Rouleau, "Le numéro spécial de la RFG-AIMS fait peau neuve ! ", *Revue française de gestion*, **1**(174), 13-14, 2007, doi: 10.3166/rfg.174.13-14.
- [13] E. MAILLOT, "Diriger un projet web Agile - Utilisez la dynamique des groupes pour décupler Scrum", *ENI*, 2015.
- [14] B. Boehm, "Spiral Model of Software Development and Enhancement", *Computer*, **21**(5), 61 - 72, 1988, doi: 10.1016/B978-0-08-051574-8.50031-5.
- [15] H. D. Benington, "Production of Large Computer Programs", *Annals of the History of Computing*, **5**(4), 350 - 361, 1983, doi: 10.1109/MAHC.1983.10102.
- [16] J. Stenbeck, "Agile project management mastery in sixty minutes, guaranteed! MI® Global Congress", *Project Management Institute*, 2010.

Stability Analysis of a DC Microgrid with Constant Power Load

Sarah Ansari*, Kamran Iqbal

University of Arkansas, Little Rock, Department of Systems Engineering, Little Rock, 72204, USA

ARTICLE INFO

Article history:

Received: 14 November, 2021

Accepted: 20 February, 2022

Online: 18 March, 2022

Keywords:

DCMG

Constant Power Load

PI Controller

Buck Converter

Cascaded Network

ABSTRACT

DC Microgrids (DCMGs) aggregate and integrate various distribution generation (DG) units through the use of power electronic converters (PECs) that are present on both the source side and the load side of the DCMGs. Tightly regulated PECs at the load side behave as constant power loads (CPLs) and may promote instability in the entire DCMG. Previous research has mostly focused on devising stabilization techniques with ideal CPLs that may not be feasible to realize; few publications that emulate DCMG stability with practical CPLs are restricted in application because they add components that considerably increase the cost of the DCMGs. This study aims at stabilizing the DCMG in the presence of practical CPL in a way that is economically feasible, i.e., without the addition of complex compensators. This paper presents a Simulink model of the smallest DCMG, i.e., a cascaded DC-DC power converter network with a practical CPL assumed at the load side of the network. Using theoretical calculations and computer simulations, we have determined the suitable CPL power level and the bandwidth of the current controller at which the smallest DCMG is stable. We have performed the stability analysis of the source side buck converter and the CPL with the derived power level and bandwidth, and found that individual converter systems are stable, thereby proving that the entire DCMG is stable despite the presence of a CPL.

1. Introduction

In recent times, the demand for DCMGs is surging. With this, there are significant issues related to the distribution networks in the power systems. The preliminary analysis and results of one such issue i.e., caused due to CPL is done in [1]. There are other associated problems like voltage fluctuations, there is a need to aggregate DG units and provide proper coordination. Thus, the development of microgrids becomes indispensable to integrate and coordinate different power systems. US department of energy, DoE defines microgrids as “Locally confined and independently controlled electric power grids in which distribution architecture integrates loads and distributed energy resources which allows the microgrid to operate connected or isolated to a main grid” [2].

While microgrids can be developed for both AC and DC supplies, DCMGs are considered superior to the AC microgrids due to several factors. The DC networks sidestep reactive power issues, which simplifies the control loop design [3]. It also results in reduced hardware (power cables), thereby reducing the overall equipment cost. Further, DCMG implementation eliminates long transmission and distribution lines that aids in providing reliable

and efficient DG system [4]. Also, in recent times, integration of renewable energy sources, fuel cells, and energy storage devices with conventional power systems has become indispensable. The urgency of these issues has brought DC power systems back into picture through DCMGs. DCMGs consist of power electronic elements that are used for various purposes. For example, they can be used to isolate the microgrid from the main power system, or to make a network of distributed generation systems that need to be synchronized. These are termed as multi-converter power electronic systems [5, 6] that employ power converters to control various grid parameters like voltage, current, power, etc.

A DC distribution system has two broad stages as shown in Figure 1 [7]. The first stage consists of two or more converters that are connected in parallel and feeding the DC bus [7]. These are switched mode power supplies (SMPS1) called line regulating converters (LRC) or source side converters. This converter system feeds the DC bus with a regulated voltage; the bus is further connected to another set of switched mode power supplies (SMPS2) called point of load (POL) converters, or load side converters. It has been shown that tightly regulated POLs behave as constant power loads (CPLs). Theoretically, the power supplied to the CPL equals the product of output voltage of the CPL and the

*Corresponding Author: Sarah Ansari, sxansari@ualr.edu

current flowing through it. When the power supplied in a CPL is constant, then the voltage varies inversely with respect to the current change. Thus, the voltage increases when the current decreases and vice versa thereby resulting in negative incremental impedance. This concludes that the constant power loads exhibit a non-linear phenomenon that causes instability in DCMGs. Moreover, solving the stability issue becomes challenging when at least two power converters are cascaded to each other. Previous literature has considered load side power converters to be behaving as ideal CPLs. As a result, study of power levels and dynamic performance of CPL and how they affect system stability has been mostly neglected. Hence, it becomes important to investigate the system stability and evaluate the technical restrictions of CPL.

This paper seeks to study what technical restrictions can be levied on CPL to ensure stability of the DCMG. To do the analysis, a simulation scheme of source side buck converter and CPL is designed. The choice of this buck topology has been reinforced by two main reasons: i) Buck topology has simple construction and dynamic performance, and ii) It has higher system stability than boost or buck-boost topology. The rest of this paper is organized as follows: sections II and III discusses the design and design of source side buck converter. Sections IV and V discusses the stability analysis and design of the CPL. Section VI discusses the cascading of the source side buck converter and CPL to form the smallest DCMG. Appendix and section VII show the simulation models developed and their corresponding results respectively. The conclusion is mentioned in section VIII.

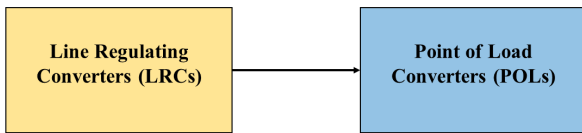


Figure 1: Major Components of DC distribution system

2. Controller Design for Source Side Buck Converter

In the research literature, linear droop control is realized through a virtual or an actual resistor in series with the DC-DC converters that are modeled as voltage sources. While droop control is a practical and viable voltage control scheme to regulate a constant DC voltage supply, it may work for buck converter topology [7]. Whereas the equivalent circuit of a converter in other topologies consists of transformers with nonlinear turn ratios, this will hinder the use of linear droop control for such converters [8]. Thus, to implement linear droop scheme, the voltage source is modeled using the DC-DC buck converter topology. Moreover, a PI controller is designed as a fast controller for the current flow through the power converters. Both controllers are proposed in this section and integrated with DC-DC buck converters to analyze their dynamic performance.

A buck converter is a power converter that steps down the DC voltage from higher input to lower output value. A buck converter with the predefined parameters is shown Figure 2. The stepping down is governed by an adjustable duty ratio which is realized by designing a suitable controller for a given buck converter.

The source side buck converter has two controllers: the voltage controller and the current controller. The load side buck converter or the CPL has a current controller and a constant power supply. This section discusses the controller design of the source side buck converter.

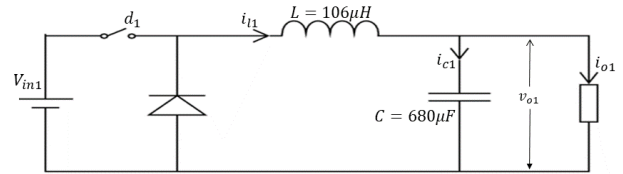


Figure 2: Schematic diagram of the source side buck converter

From the control point of view, the buck converter is considered as a power stage (shown in Figure 3), which is controlled by a PI controller. Figure 3 shows the complete control model for the buck converter with specified parameters. The objective is to design a controller to achieve the desired output voltage of 120V from an input supply of 140V. This can be done by controlling the voltage and current flowing through the power stage. Thus, the goal is to design voltage and current controllers (shown in Figure 4). In the diagram, the voltage controller compares the output voltage with the setting voltage V_{set1} (120V). Using the droop control governed by droop characteristic shown in Figure 5, the setting value of load current i_{11}^* can be obtained which is then input to the current controller. The current controller controls an inner loop consisting of a PI controller and the power stage as shown in Figure 6.

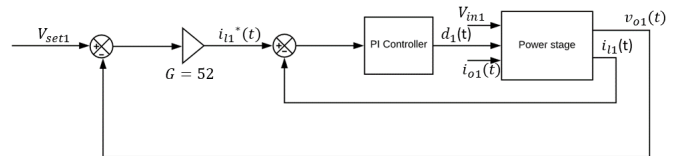


Figure 3: Control model of the source side buck converter

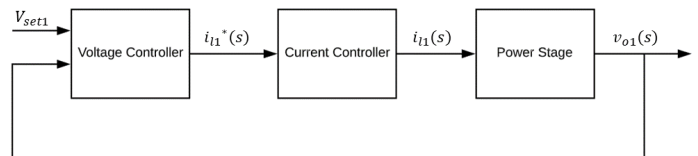


Figure 4: Voltage and current controllers for the source side buck converter

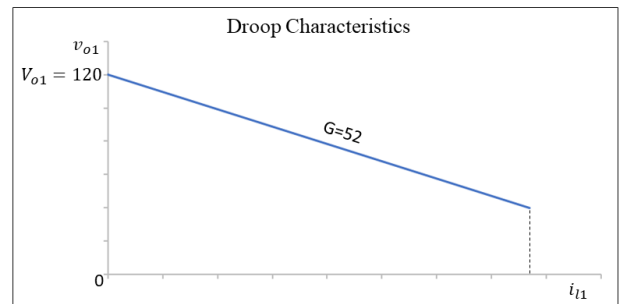


Figure 5: Droop characteristic with a droop gain of 52

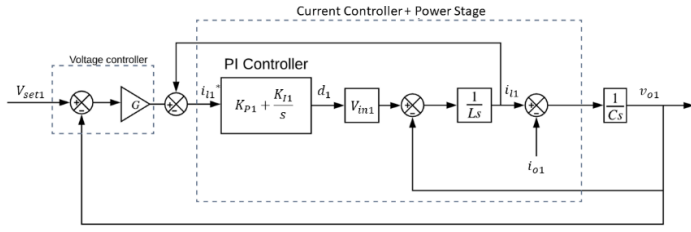


Figure 6: Closed loop current and voltage control scheme for the source side buck converter

The PI controller parameters include the proportional gain K_{P1} and the integral gain K_{I1} , which need to be determined. For design purposes, the open loop transfer function for the current control is (from Figure 6):

$$H_{ol1} = \frac{(K_{P1}s + K_{I1})V_{in1} - v_{o1}}{LS^2} \quad (1)$$

Since the dynamic change in inductor current is faster than that of the voltage across the capacitor, thus v_{o1} can be considered as a disturbance and can be neglected. Then, the modified open loop transfer function becomes

$$H_{ol1} = \frac{(K_{P1}s + K_{I1})V_{in1}}{LS^2} \quad (2)$$

The closed loop transfer function comprising of PI controller, power stage and unity feedback is

$$H_{cl1} = \frac{H_{ol1}}{1 + H_{ol1}} = \frac{(K_{P1}s + K_{I1})}{(L/V_{in1})s^2 + K_{P1}s + K_{I1}} \quad (3)$$

3. Stability Analysis of Source Side Converter

The model of the source side buck converter was implemented in Simulink. The stability analysis of the source side buck converter was undertaken in two distinct ways, as described below:

Approach 1: the current loop design

In this approach the current loop is analyzed and the current controller consisting of the power stage and the PI controller is considered. The complete model is shown in Figure 7.

The transfer function of the PI controller $G1_{c1}(s)$ is:

$$G1_{c1}(s) = K_{P1} + \frac{K_{I1}}{s} = K_{P1} \left(\frac{s + (K_{I1}/K_{P1})}{s} \right) \quad (4)$$

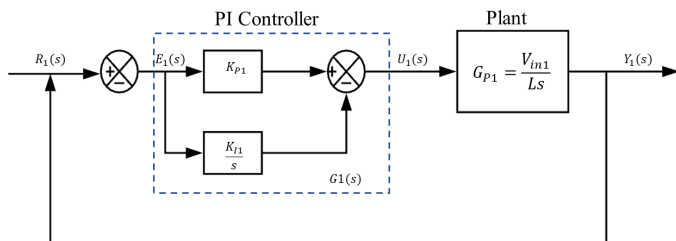


Figure 7: Control system implementation of the source side buck converter

The open-loop transfer function is

$$G1_{ol1}(s) = G1_{P1}(s)G1_{c1}(s) = \frac{V_{in1}K_{P1}(s + (K_{I1}/K_{P1}))}{LS^2} \quad (5)$$

Let $K_{I1}/K_{P1} = K1$, then Equation (5) becomes

$$G1_{ol1}(s) = \frac{V_{in1}K_{P1}(s + K1)}{LS^2} \quad (6)$$

The characteristic equation of the closed-loop system is given as $1 + G1_{ol1}(s)$, where

$$s^2 + \frac{V_{in1}K_{P1}}{L}s + \frac{V_{in1}K1}{L} = 0 \quad (7)$$

The closed-loop transfer function is

$$G1_{cl1} = \frac{Y1(s)}{R1(s)} = \frac{(V_{in1}/L)(K_{P1}s + K1)}{s^2 + (V_{in1}K_{P1}/L)s + (V_{in1}K1/L)} \quad (8)$$

Approach 2: voltage and current loop design

In this approach, the entire system including voltage and current controllers is considered. The controller structure as shown in Figure 6 will be considered for the analysis.

The open-loop transfer function system for the design of PI controller is derived from Matlab:

$$G1_{c2}(s) = \frac{1.32 \times 10^6 s + 1.011 \times 10^{11}}{s^2 + 1.387 \times 10^7} \quad (9)$$

The loop transfer function with PI controller in the loop is (where, $K1 = K_{I1}/K_{P1}$)

$$G1_{ol2}(s) = \frac{1.32 \times 10^6 s + 1.011 \times 10^{11}}{s^2 + 1.387 \times 10^7} \times \left(\frac{K_{P1}(s + K1)}{s} \right) \quad (10)$$

The stability analysis and design of both approaches is performed using Root locus method, Routh Hurwitz criterion, and Nyquist criterion.

3.1. Root Locus Method

The root locus-based design aims to find suitable values for the proportional and integral gains.

Approach 1: The open-loop transfer function (Equation 6) of the current controller is studied by varying gain K_{P1} . The characteristic equation of (6) is

$$s^2 + \frac{V_{in1}K_{P1}}{L}s + \frac{V_{in1}K1}{L} = 0 \quad (11)$$

The resulting root loci of Equation (5) with $K_{I1}/K_{P1} = 2 \times 10^5$ are shown in Figure 8. From the root loci plot, when gain $K_{P1} = 0.568$, the damping ratio is 0.968, which is considered reasonable for the converter. The corresponding value of $K_{I1} = 113600$. The characteristic equation has two complex roots (also shown in Figure 8) at:

$$s = -3.75 \times 10^5 + j9.69 \times 10^4 \quad \text{and} \quad s = -3.75 \times 10^5 - j9.69 \times 10^4 \quad (12)$$

These are the poles of the closed loop system. Since, these poles are located in the LHP, the closed loop system is stable with reasonable damping.

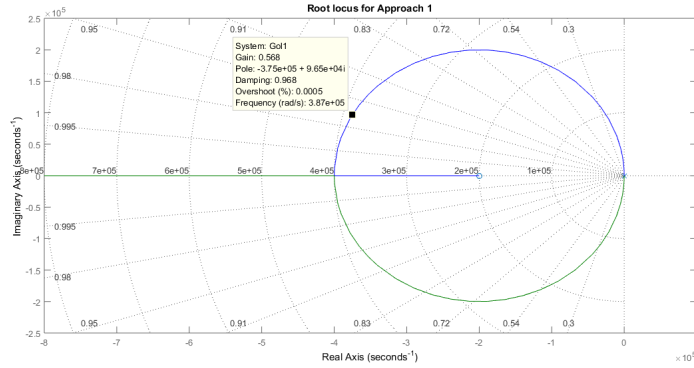


Figure 8: Root Loci of equation (11) with $K_{I1}/K_{P1} = 2 \times 10^5$; K_{P1} varies

Approach 2: In this case, similarly, the gain ratio $K_{I1}/K_{P1} = 2 \times 10^5$ is considered. Equation (10) gives the open-loop transfer function of the system with voltage and current controllers. The root loci of (10) are shown in Figure 9. Clearly, the poles of the closed-loop system are located in the LHP. Thus, the closed loop system is stable.

The closed loop transfer function with $K_{P1} = 0.568$ and $K_{I1} = 113600$ is

$$G1_{cl2}(s) = \frac{7.498 \times 10^5 s^2 + 2.073 \times 10^{11} s + 1.147 \times 10^{16}}{s^3 + 7.498 \times 10^5 s^2 + 2.073 \times 10^{11} s + 1.147 \times 10^{16}} \quad (13)$$

Thus, the closed loop system is stable.

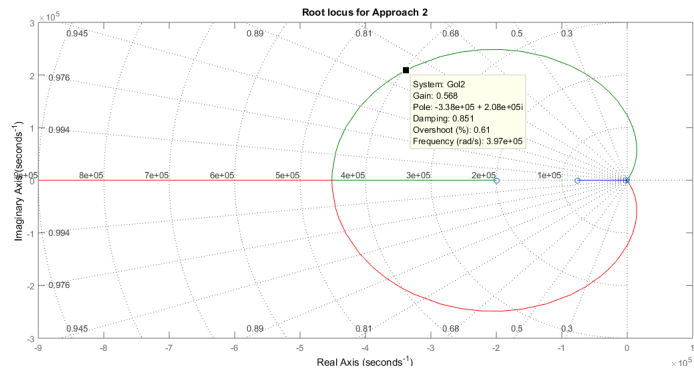


Figure 9: Root Loci of equation (10) with $K_{I2}/K_{P2} = 2 \times 10^5$; K_{P2} varies

3.2. Routh-Hurwitz Criterion

The Routh-Hurwitz criterion algebraically ascertains the stability requirements for a linear time-invariant (LTI) system modeled with constant coefficients. The criterion tests whether any roots of the characteristic equation lie in the right half s -plane.

Approach 1: The characteristic equation of the closed-loop system is given as $1 + G1_{ol1}(s)$ and is

$$s^2 + \frac{V_{in1} K_{P1}}{L} s + \frac{V_{in1} K_{I1}}{L} = 0 \quad (14)$$

Applying the Routh Hurwitz's stability criterion to equation (7) yields that the system is stable for $K_{P1} > 0$ and $K_{I1} > 0$. Thus, the chosen parameter values of $K_{P1} = 0.568$ and $K_{I1} = 113600$ stabilize the system.

Approach 2: The characteristic equation of the closed-loop system is

$$s^3 + 7.5 \times 10^5 s^2 + 2.1 \times 10^{11} s + 1.1 \times 10^{16} = 0 \quad (15)$$

Then, for the closed-loop system to be stable, it should meet the following constraints:

$$7.5 \times 10^5 \times 2.1 \times 10^{11} > 1.1 \times 10^{16}$$

$$1.5 \times 10^{17} > 1.1 \times 10^{16} \quad (16)$$

The above design satisfies these constraints; hence, the system is stable.

3.3. Nyquist Criterion

The Nyquist criterion graphically reveals information about the number of poles and zeroes of the closed-loop transfer function that are in the right half s -plane. The Nyquist criterion is applied to the two design approaches as follows.

Approach 1: The Nyquist plot of the open loop transfer function $G1_{ol1}(s)$ (from equation (6)) with $K_{P1} = 0.568$ and $K_{I1} = 113600$ is shown in Figure 10, where

$$G1_{ol1}(s) = \frac{79.52s + 1.59 \times 10^7}{106 \times 10^{-6} s^2} \quad (17)$$

For a minimum phase transfer function, the closed-loop system is stable if there are no encirclements of the critical point $(-1 + j0)$. From Figure 10, since there are no encirclements of the critical point, thus the closed-loop system is stable. This result is also verified by Matlab.

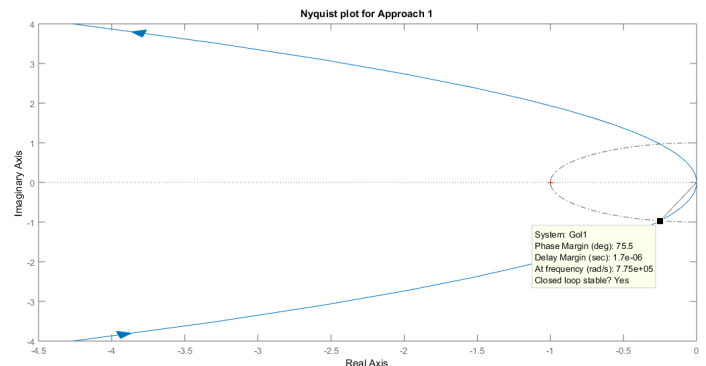


Figure 10: Nyquist plot for Approach 1

Approach 2: The Nyquist plot of the open loop transfer function $G_{1_{ol2}}(s)$ (from (10)) with $K_{P1} = 0.568$ and $K_{I1} = 113600$ is shown in Figure 11, where the loop transfer function is given as

$$G_{1_{ol2}}(s) = \frac{7.498 \times 10^5 s^2 + 2.073 \times 10^{11} s + 1.147 \times 10^{16}}{s^3 + 1.387 \times 10^7 s} \quad (18)$$

From Figure 11, there are no encirclements of the critical point, hence the closed-loop system is stable. This result is also verified by Matlab.

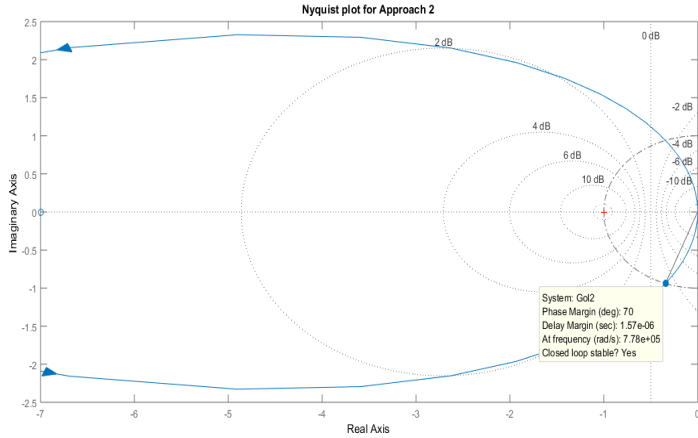


Figure 11: Nyquist plot for Approach 2

Based on the above stability criteria, the values of K_{P1} and K_{I1} (reported in Table 1 below) stabilize the source side buck converter. The simulation model and corresponding results are shown in appendix and section VII respectively.

Table 1: Values of K_{P1} and K_{I1} for the given source side buck converter

K_{P1}	0.568
K_{I1}	113600

4. Design of CPL

Buck converters can be emulated as instantaneous constant power loads when cascaded with at least one source side DC-DC power converters. For the study, one source side buck converter is considered, and its controller design is proposed in the previous section. Figure 12 shows the control model for a power stage (here buck converter) emulated as CPL.

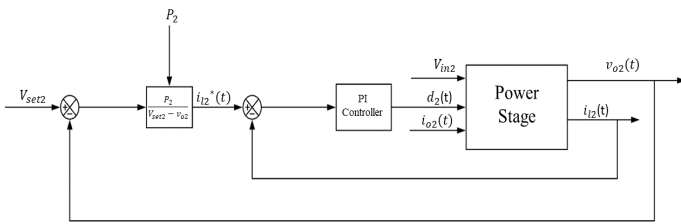


Figure 12: Control Model of CPL

Also, the power stage is supplied with a constant supply of power P_2 which characterizes the non-linear nature of the CPL. It thus becomes important to linearize the system about an equilibrium point.

4.1. Linearization of Load Side Converter (CPL)

From figure 12 the relationship between setting value of inductor current i_{L2}^* and incoming voltage of the CPL is non-linear and is given as,

$$i_{L2}^* = \frac{P_2}{v_{o2}} \quad (19)$$

Here, V_{set2} is not considered because the component is added to the simulation model to the closed loop system. Theoretical analysis of the CPL that involves linearization of CPL and its related calculation is based on the open loop circuitry of the CPL which does not have V_{set2} . Each of the parameters in Figure 12 can also be represented as the sum of steady state value at equilibrium point and the small signal perturbation around the equilibrium as shown in equation 20.

$$\left. \begin{aligned} i_{L2}^* &= I_{L2}^* + \tilde{i}_{L2}^* \\ v_{o2} &= V_{o2} + \tilde{v}_{o2} \\ P_2 &= P_2 + \tilde{p}_2 \end{aligned} \right\} \quad (20)$$

Using Taylor series as explained in [9], the linearized equation is

$$i_{L2}^* = \frac{1}{V_{o2}} \tilde{p}_2 - \frac{P_2}{V_{o2}^2} \tilde{v}_{o2} \quad (21)$$

Hence, the linearized model of CPL is shown in Figure 13.

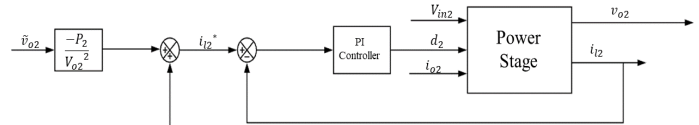


Figure 13: Linearized version of CPL

4.2. PI Control for Linearized CPL (Load Side Buck Converter)

The PI controller parameters namely proportional gain K_{P2} and integral gain K_{I2} need to be determined. The open loop transfer function for the current control is (from Figure 14):

$$H_{ol2} = \frac{(K_{P2}s + K_{I2})V_{in2} - v_{o2}}{LS^2} \quad (22)$$

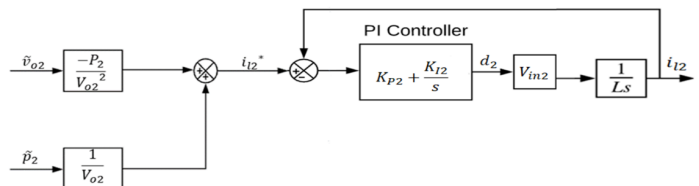


Figure 14: PI control for linearized CPL

Since the dynamic change in inductor current is faster than that of the voltage across the capacitor, thus v_{o2} can be considered

as a disturbance and can be neglected. Thus, the modified open loop transfer function becomes

$$H_{ol2} = \frac{(K_{P2}s + K_{I2})V_{in2}}{LS^2} \quad (23)$$

The closed loop transfer function comprising of PI controller, power stage and unity feedback is

$$H_{cl2} = \frac{H_{ol2}}{1 + H_{ol2}} = \frac{(K_{P2}s + k_{I2})}{(L/V_{in2})s^2 + K_{P2}s + K_{I2}} \quad (24)$$

Equation (24) is the closed loop transfer function of the current controller which is the PI controller and the power stage.

4.3. Power and Bandwidth of CPL

The power stage of the CPL used in this study is the same as that of the source side buck converter. The function of the CPL is to step down the voltage from 120V to 100V. In order to design a PI controller for such a CPL, we have assumed the values of $K_{P2} = K_{P1}$ and $K_{I2} = K_{I1}$. This is done, due to two reasons:

1. In practical DCMGs, it becomes favorable to have similar current controllers for the source side converter and CPL, as it reduces the complexity of the cascaded network.
2. By doing so, the dynamic behavior of both the converters can be compared in order to better understand the working of the CPL.

Thus, $K_{P2} = 0.568$ and $K_{I2} = 113600$.

Since in a CPL, the power supplied is constant, thus $\tilde{p}_2 = 0$. Thus, equation (21) is modified and is given as,

$$i_{l2}^* = -\frac{P_2}{V_{o2}^2} \tilde{v}_{o2} \quad (25)$$

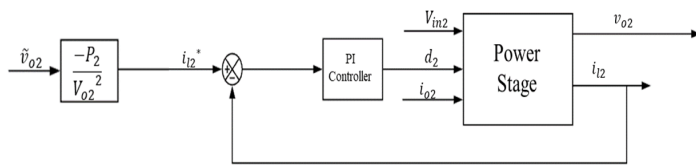


Figure 15: Linearized CPL, ignoring $\frac{1}{V_{o2}} \tilde{p}_2$, since $\tilde{p}_2 = 0$

Notice, in Figure 15, the value of the incremental impedance is positive as now the CPL emulates a resistive load, thereby ensuring stability to the DCMG, by keeping its property intact.

Now,

$$I_{l2}^* = \frac{P_2}{V_{o2}^2} \quad (26)$$

Since, the parameters of the source side buck converter and that of the CPL is considered the same, thus the droop control of the

source side buck converter is analogous to the $\frac{P_2}{V_{o2}^2}$ factor. Thus, assuming $\frac{P_2}{V_{o2}^2} = G = 52$ and V_{o2} is the desired output voltage of CPL, which is 100V, thus, we get

$$P_2 = 52 \times V_{o2}^2 = 52 \times (100)^2 = 520kW \quad (27)$$

Using this value of power (derived in Equation (26)), we have done the stability analysis of the CPL to verify that at $P_2 = 520kW$ the CPL is stable.

5. Stability Analysis of CPL (Load Side Converter)

Stability analysis of the Simulink model of the buck converter is similarly done in two distinct ways, as described below:

Approach 1: the current loop design

In this approach the current controller consisting of the power stage and PI controller is considered. The complete model is shown in Figure 16. The transfer function of the PI controller $G2_{c1}(s)$ is:

$$G2_{c1}(s) = K_{P2} + \frac{K_{I2}}{s} = K_{P2} \left(\frac{s + (K_{I2}/K_{P2})}{s} \right) \quad (28)$$

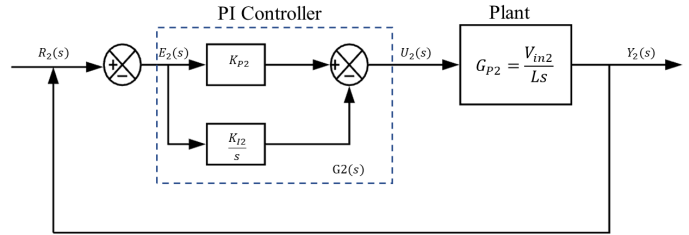


Figure 16: Control system of the CPL

The forward-path transfer function of the feedback control system is

$$G2_{ol1}(s) = G2_{P2}(s)G2_{c1}(s) = \frac{V_{in2}K_{P2}(s + (K_{I2}/K_{P2}))}{LS^2} \quad (29)$$

Let $K_{I2}/K_{P2} = K2$, thus Eq (29) becomes

$$G2_{ol1}(s) = \frac{V_{in2}K_{P2}(s + K2)}{LS^2} \quad (30)$$

The characteristic equation of the closed-loop system as given by $1 + G2_{ol1}(s)$ is

$$s^2 + \frac{V_{in2}K_{P2}}{L}s + \frac{V_{in2}K_{I2}}{L} = 0 \quad (31)$$

The closed-loop function is

$$G2_{cl1} = \frac{Y_2(s)}{R_2(s)} = \frac{(V_{in2}/L)(K_{P2}s + K_{I2})}{s^2 + (V_{in2}K_{P2}/L)s + (V_{in2}K_{I2}/L)} \quad (32)$$

Approach 2: Current and voltage loop designs

In this approach, the entire system with linearized CPL (having $\frac{P_2}{V_{o2}}$) and current controller (shown in Figure 15) is considered. The open loop transfer function system with analysis point as the PI controller, is derived from Matlab:

$$G2_{c2}(s) = \frac{7.208 \times 10^{-4} s^3 + 979.2 s^2 + 5.396 \times 10^9 s + 1.04 \times 10^{15}}{7.208 \times 10^{-4} s^3 + s} \quad (33)$$

The open loop transfer function of the system with PI controller is given below, where $K2 = K_{I2}/K_{P2}$

$$G2_{ol2}(s) = \frac{7.208 \times 10^{-4} s^3 + 979.2 s^2 + 5.396 \times 10^9 s + 1.04 \times 10^{15}}{7.208 \times 10^{-4} s^3 + s} \times \left(\frac{K_{P2}(s + K2)}{s} \right) \quad (34)$$

Three stability analysis criteria are employed toward controller design. These include: The Root Locus method, the Routh Hurwitz criterion and the Nyquist criterion.

5.1. Root Locus Method

The root locus method is aimed to find suitable values for the proportional and integral gains.

Approach 1: Equation (30) gives the open loop transfer function of the current controller with varying gain K_{P2} . The closed-loop characteristic equation for (30) is

$$s^2 + \frac{V_{in2} K_{P2}}{L} s + \frac{V_{in2} K_{I2}}{L} = 0 \quad (35)$$

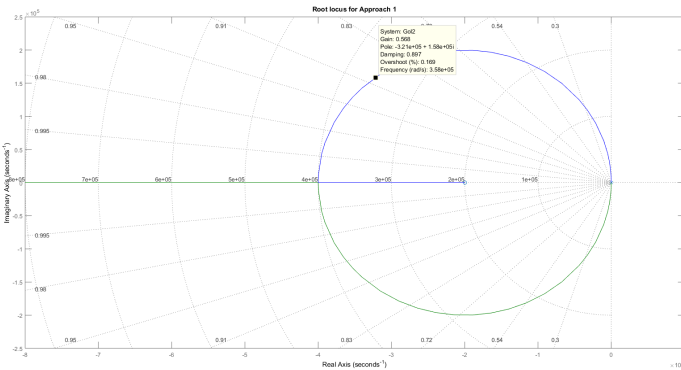


Figure 17: Root Loci of equation (35) with $K_{I2}/K_{P2} = 2 \times 10^5$; K_{P2} varies

The root loci of (30) with $K_{I2}/K_{P2} = 2 \times 10^5$ are shown in Figure 17. From the root loci, when gain $K_{P2} = 0.568$, the damping ratio is 0.897, which is considered reasonable for the converter. Consequently, the value of $K_{I2} = 113600$ is selected. The two characteristic equation roots are (shown in Figure 17) are at

$$s_1 = -3.21 \times 10^5 + j1.59 \times 10^5 \quad \text{and} \quad s_2 = -3.21 \times 10^5 - j1.59 \times 10^5 \quad (36)$$

Since these closed-loop poles lie in the LHP, the closed loop system is stable.

Approach 2: In this case, similarly, the design ratio $K_{I2}/K_{P2} = 2 \times 10^5$ is considered. Equation (34) gives the open loop transfer function of the linearized CPL with the current controller. The root loci of (34) are shown in Figure 18. From the graph, the poles of the closed-loop system lie on the LHP. The closed loop transfer function with $K_{P2} = 0.568$ and $K_{I2} = 113600$ is

$$G2_{cl2}(s) = \frac{4.094 \times 10^{-4} s^4 + 638.1 s^3 + 3.176 \times 10^9 s^2 + 1.204 \times 10^{15} s + 1.181 \times 10^{20}}{1.13 \times 10^{-3} s^4 + 638.1 s^3 + 3.176 \times 10^9 s^2 + 1.204 \times 10^{15} s + 1.181 \times 10^{20}} \quad (37)$$

Thus, the closed loop system is stable.

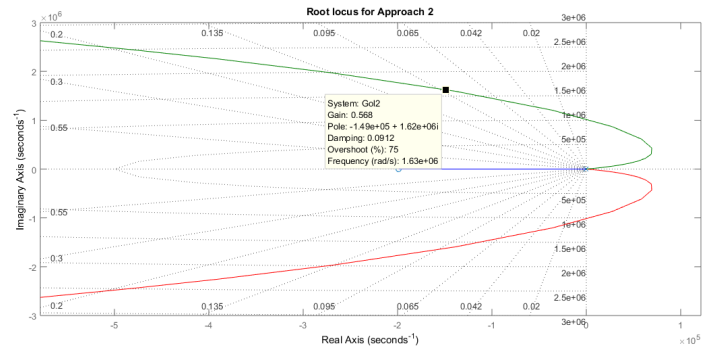


Figure 18: Root Loci of equation (34) with $K_{I2}/K_{P2} = 2 \times 10^5$; K_{P2} varies

5.2. Routh-Hurwitz Criterion

Routh-Hurwitz criterion is similarly applied to determine the conditions for closed-loop stability.

Approach 1: The characteristic equation of the closed-loop system, given as $1 + G2_{ol1}(s)$, is

$$s^2 + \frac{V_{in2} K_{P2}}{L} s + \frac{V_{in2} K_{I2}}{L} = 0 \quad (38)$$

Applying the Routh Hurwitz's stability criterion to equation (33) yields the result that the system is stable for $K_{P2} > 0$ and $K_{I2} > 0$. Thus, for the chosen values of $K_{P2} = 0.568$ and $K_{I2} = 113600$, the closed-loop system is stable.

Approach 2: The characteristic equation of the closed-loop system is

$$1.1 \times 10^{-3} s^4 + 638.1 s^3 + 3.2 \times 10^9 s^2 + 1.2 \times 10^{15} s + 1.2 \times 10^{20} = 0 \quad (39)$$

For the system to be stable, each of the diagonal minors ($\Delta_1, \Delta_2, \Delta_3$) should be zero, i.e.,

$$\begin{aligned} \Delta_1 &= 638.1 > 0 \\ \Delta_2 &= 7.2192 \times 10^{11} > 0 \\ \Delta_3 &= 8.16 \times 10^{26} > 0 \end{aligned} \quad (40)$$

From the conditions in equation (40), it is evident that for the selected parameter values, each of the diagonal minors are greater than 0, thus the closed-loop system is stable.

5.3. Nyquist Criterion

The Nyquist criterion is similarly applied to ascertain the stability of the closed-loop system.

Approach 1: The Nyquist plot of the open loop transfer function $G2_{ol1}(s)$ (from equation (30)) with $K_{P2} = 0.568$ and $K_{I2} = 113600$ is given as

$$G2_{ol1}(s) = \frac{79.52s + 1.59 \times 10^7}{106 \times 10^{-6} s^2} \quad (41)$$

For a minimum phase transfer function, the closed-loop system is stable if there no encirclements of the critical point $(-1 + j0)$. Since, from Figure 19, there are no encirclements of the critical point, thus the closed loop system is stable. It is also verified by Matlab.

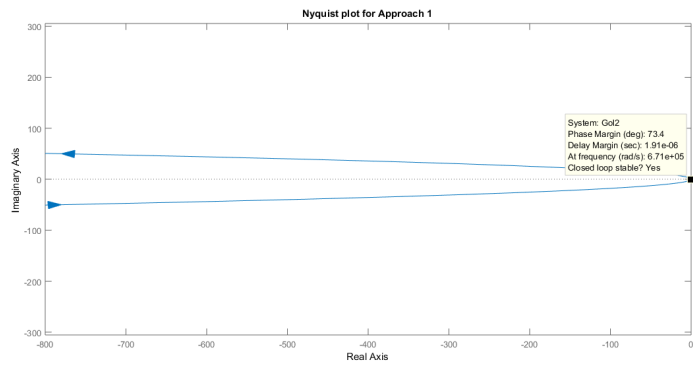


Figure 19: Nyquist plot for Approach 1

Approach 2: The Nyquist plot of the open loop transfer function $G2_{ol2}(s)$ (from (34)) with $K_{P2} = 0.568$ and $K_{21} = 113600$ is shown in Figure 20, where the loop transfer function is given as

$$G2_{ol2}(s) = \frac{4.094 \times 10^{-4} s^4 + 638.1s^3 + 3.176 \times 10^9 s^2 + 1.204 \times 10^{15} s + 1.181 \times 10^{20}}{7.208 \times 10^{-4} s^4 + s^2} \quad (42)$$

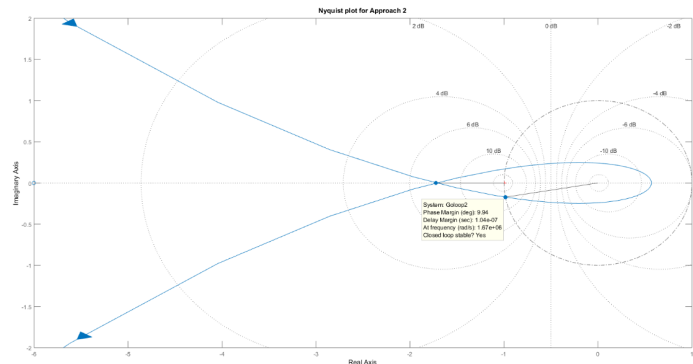


Figure 20: Nyquist plot for Approach 2

The Approach 1 takes into consideration only the current controller and the power stage. The resulting bandwidth BW_2 of the closed loop transfer function is the bandwidth when the CPL

is stable. The values of P_2 and BW_2 are shown below in Table 2. The simulation model of CPL is shown in Appendix and the results are shown in section VII.

Table 2: Derived values of P_2 and BW_2 for the CPL

P_2	520kW
BW_2	8.321×10^5 rad/sec

6. Cascaded Network of Source Side Converter and CPL

Cascading two power converters means that the load side converter behaves as the load of the source side converter. In other words, the output voltage of source side converter acts as the input voltage of load side converter, and the output inductor current of source side converter is fed as an input to the load side converter [10].

In our study, we have cascaded the source side buck converter with the linearized CPL. Cascading can be done only when both the power converters are independently stable. In our study, we have shown that the source side buck converter and the CPL are independently stable. Thus, they can be cascaded, thereby forming the smallest DCMG.

In our study, the output voltage of the source side buck converter, V_{o1} , is fed as the input voltage to the CPL as V_{in2} . Also, the output inductor current of the source side buck converter, I_{L1} is fed as input current to the CPL, I_{o2} .

Table 3 gives the values of all the parameters of the cascaded power converters, also considered as the smallest DCMG. The simulation model of the cascaded network is shown in Appendix and the results are shown in Sections VII.

Table 3: Values of all the parameters of cascaded source side buck converter and CPL

V_{in1}	140V
I_{o1}	20A
V_{o1}	120V
I_{L1}	20A
$V_{in2} = V_{o1}$	120V
$I_{o2} = I_{L1}$	20A
V_{o2}	100V
I_{L2}	20A
P_2	520kW
BW_2	8.321×10^5 rad/sec
L	106μH
C	680μF

7. Simulation Results of DCMG

The Simulink model for the DCMG is shown in Figure 27 (see Appendix). When the model is simulated, the inductor currents of the source side buck converter and CPL will charge their respective capacitors in order to increase the voltage across the capacitor from 0 to 120V (in case of source buck converter) and from 0 to 100V (in case of CPL). When the load current of 20A is supplied, the voltage across the capacitor decreases slightly, resulting in the inductor current exceeding 20A. The output voltage of 120 V (shown in Figure 21) from the source side buck

converter is fed as an input voltage to the CPL. The resulting output voltage of the CPL is 100 V (shown in Figure 23). The output inductor current of 20A (shown in Figure 22) from the source side buck converter is fed as an input current to the CPL. The resulting output inductor current of the CPL also comes out to be 20A (shown in Figure 24). This confirms the cascading of the source side converter with the CPL, to form the smallest DCMG. Clearly, there is no overshoot and no oscillations observed in Figures 21-24.

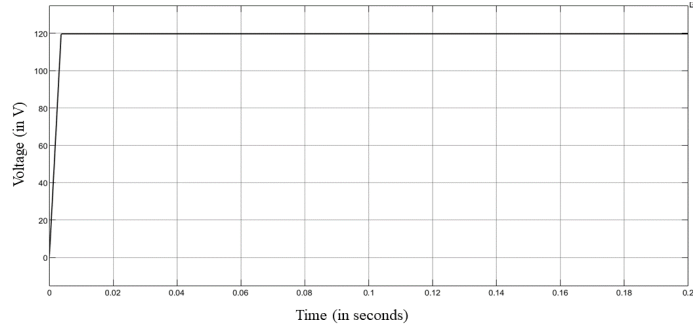


Figure 21: Oscilloscope result of output voltage of source side buck converter

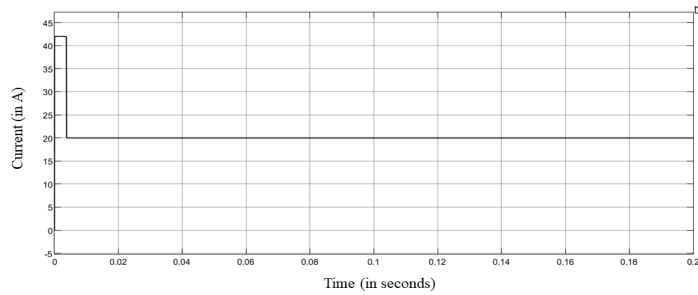


Figure 22: Oscilloscope result of output inductor current of source side buck converter

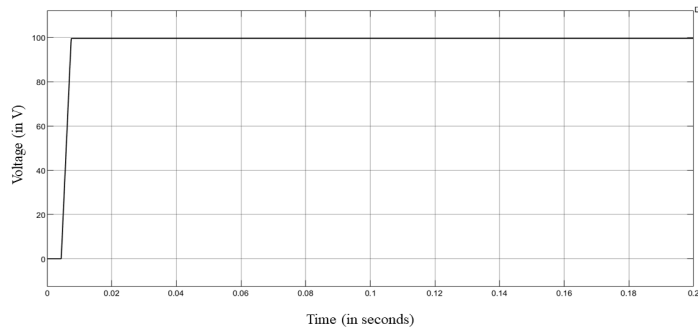


Figure 23: Oscilloscope result of output voltage of CPL

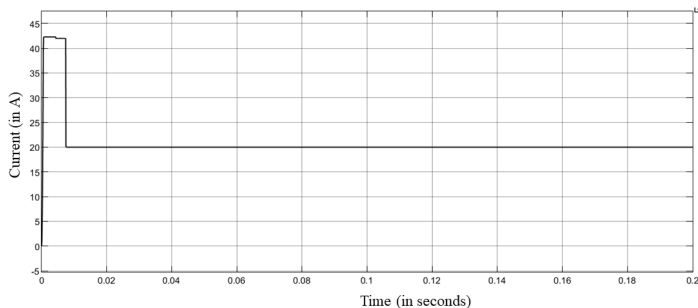


Figure 24: Oscilloscope result of inductor current of CPL

8. Conclusion

This paper discussed the stability analysis of the smallest DCMG that consists of a source side buck converter and a CPL. The cascaded power converters are abundantly found in the DCMGs, and power converters located at the load side that act as CPLs have the potential to cause instability to the entire DCMGs. Thus, it is important to eliminate the instability caused due to CPLs, so that the entire DCMG is stable. Keeping this in mind, the stability analysis of cascaded DC-DC power converters was done. This research proved that despite the presence of CPL, the DCMG can still be made stable. The stability analysis done on the individual components of the cascaded network draws interesting conclusions that support the fact that the DCMG can be stable at certain power level and bandwidth of the CPL controller.

The following are the main results that can be drawn from this research:

1. The CPL, that causes instability to the entire DCMG is stabilized at a power level of 520kW and bandwidth of $8.321 \times 10^5 rad/sec$.
2. The individual components of the cascaded network, consisting of source side converter and CPL (load side converter) are stable in steady state, thereby making the DCMG stable.
3. The DCMG consisting of CPL in cascaded DC-DC power converter network is stable at a certain power level of the load. The power of the load is found out to be 520kW.
4. The DCMG consisting of CPL in cascaded DC-DC power converter network is stable with controller bandwidth of $8.321 \times 10^5 rad/sec$, which is the bandwidth of the current controller of the CPL.

It is important to note that the stability analysis of the DCMG with CPL is done with specific parameter values used in this study. The stability analysis can be repeated with a different set of controller design parameters.

Appendix

This paper describes the design and stability analysis of:

1. Source side buck converter.
2. CPL (emulated as buck converter) and
3. Cascaded network of source side buck converter and CPL.

In order to verify that the theoretical results and calculations align well, we have simulated the models using Simulink software. The result of the cascaded network is shown in section VII.

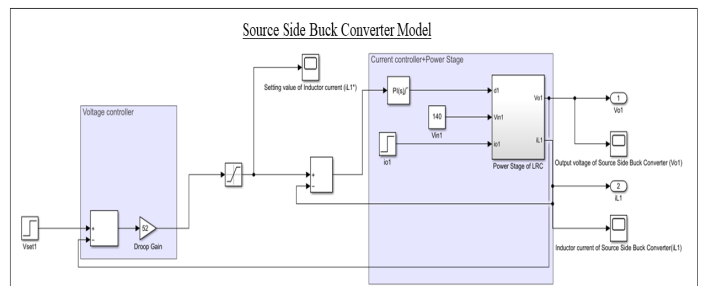


Figure 25: Simulink Model of Source Side Buck Converter

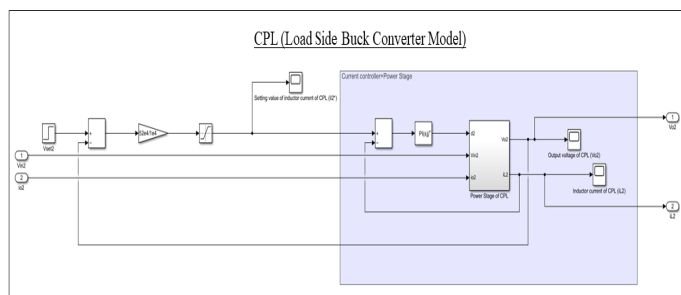


Figure 26: Simulink Model of CPL

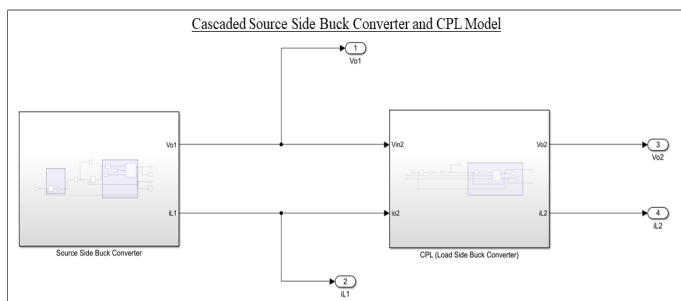


Figure 27: Simulink Model of the cascaded network of source side buck converter and CPL

stability analysis and experimental verification." IET Power Electronics **11**(9), 1519-1528, 2018, doi: 10.1049/iet-pe.2017.0670.

- [10] M. Cupelli, L. Zhu, A. Monti, "Why ideal constant power loads are not the worst case condition from a control standpoint." IEEE Transactions on Smart Grid **6**(6), 2596-2606, 2014, doi: 10.1109/TSG.2014.2361630.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors are grateful to the Department of Systems Engineering, University of Arkansas, Little Rock, USA for providing necessary facilities to perform the experiments.

References

- [1] S. Ansari, J. Zhang, K. Iqbal, "Modeling, Stability Analysis and Simulation of Buck Converter in a DC Microgrid," in 2021 IEEE Kansas Power and Energy Conference (KPEC), 1-4, 2021, doi: 10.1109/KPEC51835.2021.9446255.
- [2] T. Dragičević, X. Lu, J. C. Vasquez, J. M. Guerrero, "DC microgrids— Part I: A review of control strategies and stabilization techniques", IEEE Trans. Power Electron., **31**(7) 4876-4891, 2016, doi: 10.1109/TPEL.2015.2478859.
- [3] D. Olson, "Current market electricity supply issues & trends. It is all about the Peak," in 2008 IEEE Intl. Telecom. Energy Confer. (INTELEC), 1-5, 2008, doi: 10.1109/INTLEC.2008.4664017.
- [4] S. Luo, "A review of distributed power systems part I: DC distributed power system" IEEE Aerosp. Electron. Syst. Mag. **20**(8), 5-16, 2005, doi: 10.1109/MAES.2005.1499272.
- [5] D.K. Fulwani, S. Singh, "Mitigation of Negative Impedance Instabilities in DC Distribution Systems: A Sliding Mode Control Approach", Springer: Berlin, 2016.
- [6] V. Grigore, J. Hatonen, J. Kyyra, T. Suntio, "Dynamics of a buck converter with a constant power load," in IEEE Power Electron. Spec. Conf. (PESC 98), vol. 1, 72-78, 1998, doi: 10.1109/PESC.1998.701881.
- [7] M. Srinivasan, "Hierarchical Control of DC Microgrids with Constant Power Loads," Ph.D. Thesis, The University of Texas, 2017.
- [8] R.W. Erickson and D. Maksimovic, Fundamentals of Power Electronics Norwell, MA: Kluwer, 1997.
- [9] R. Gavagsaz-Ghoachani, L.-M. Saublet, M. Phattanasak, J.-P. Martin, B. Nahid-Mobarakeh, S. Pierfederici, "Active stabilisation design of DC-DC converters with constant power load using a sampled discrete-time model:

On the Construction of Symmetries and Retaining Lifted Representations in Dynamic Probabilistic Relational Models

Nils Finke*, Ralf Möller

Institute of Information Systems, Universität zu Lübeck, 23562 Lübeck, Germany

ARTICLE INFO

Article history:

Received: 11 January, 2022

Accepted: 05 March, 2022

Online: 18 March, 2022

Keywords:

Relational Models

Lifting

Ordinal Pattern

Symmetry

ABSTRACT

Our world is characterised by uncertainty and complex, relational structures that carry temporal information, yielding large dynamic probabilistic relational models at the centre of many applications. We consider an example from logistics in which the transportation of cargoes using vessels (objects) driven by the amount of supply and the potential to generate revenue (relational) changes over time (temporal or dynamic). If a model includes only a few objects, the model is still considerably small, but once including more objects, i.e., with increasing domain size, the complexity of the model increases. However, with an increase in the domain size, the likelihood of keeping redundant information in the model also increases. In the research field of lifted probabilistic inference, redundant information is referred to as symmetries, which, informally speaking, are exploited in query answering by using one object from a group of symmetrical objects as a representative in order to reduce computational complexity. In existing research, lifted graphical models are assumed to already contain symmetries, which do not need to be constructed in the first place. To the best of our knowledge, we are the first to propose symmetry construction a priori through a symbolisation scheme to approximate temporal symmetries, i.e., objects that tend to behave the same over time. Even if groups of objects show symmetrical behaviour in the long term, temporal deviations in the behaviour of objects that are actually considered symmetrical can lead to splitting a symmetry group, which is called grounding. A split requires to treat objects individually from that point on, which affects the efficiency in answering queries. According to the open-world assumption, we use symmetry groups to prevent groundings whenever objects deviate in behaviour, either due to missing or contrary observations.

1 Introduction

This paper is an extension of two works originally presented in *KI 2021: Advances in Artificial Intelligence* [1] and in *AI 2021: Advances in Artificial Intelligence – 34rd Australasian Joint Conference* [2]. Both papers study the approximation of symmetries using an ordinal pattern symbolisation approach to prevent groundings in dynamic probabilistic relational models (DPRMs)¹.

In order to cope with uncertainty and relational information of numerous objects over time, in many real-world applications, probabilistic temporal (also called dynamic) relational models (DPRMs) need often be employed [3]. Reasoning on large probabilistic models, like in data-driven decision making, often requires evaluating multiple scenarios by answering sets of queries, e.g., regarding the probability of events, probability distributions, or actions leading to a maximum expected utility (MEU). Further, reasoning on large

probabilistic models is often performed under time-critical conditions, i.e., where computational tractability is essential [4]. In this respect, DPRMs, together with lifted inference approaches, provide an efficient formalism addressing this problem. DPRMs describe dependencies between objects, their attributes and their relations in a sparse manner. To encode uncertainty, DPRMs encode probability distributions by exploiting in-dependencies between random variables (randvars) using factor graphs. Factor graphs are combined with relational logic, using logical variables (logvars) as parameters for randvars to compactly represent sets of randvars that are considered indistinguishable (without further evidence). This technique is also known as *lifting* [5, 6]. A lifted representation of a probabilistic graphical model allows for a sparse representation to restrain state complexity and enables to decrease runtime complexity in inference.

To illustrate the potential of lifting, let us think of creating a probabilistic model for navigational route planning and congestion

*Corresponding Author: Nils Finke, Institute of Information Systems, Universität zu Lübeck, 23562 Lübeck, Germany, finke@ifis.uni-luebeck.de

¹pronounced *deeper* models

avoidance in dry-bulk shipping. Dry-bulk shipping is one of the most important forms of transportation as part of the global supply chain [7, 3]. Especially the last year 2020, which was marked by the coronavirus pandemic, shows the importance of good supply chain management. An important sub-challenge in supply chain management is congestion avoidance, which has been studied in research ever-since [8]-[10]. Setting up a probabilistic model to improve planning and to avoid congestion requires identifying features, such as demand for commodities and traffic volume, affecting any routing plans. Commodities are unevenly spread across the globe due to the different mineral resources of countries. In case of excessive demand, regions where the demanded commodities are mined and supplied are excessively visited for shipping, resulting in congestion in those regions. If such a model includes only a few objects, here regions from which commodities are transported, the model might still be considerably small, but once including more regions to capture the whole market, i.e., with an increasing domain size, the complexity of the model increases. However, with an increase in complexity due to an increase in the domain size also the likelihood of keeping redundant information within the model increases. For example, in the application of route planning, multiple regions may exist that are similar in terms of features of the model. Intuitive examples are regions offering the same commodities, i.e., regions with similar mineral resources.

Lifting exactly exploits that existence: Regions which are symmetrical with respect to the features used in the model can be treated by one representative for a group of symmetrical objects to obtain a sparser representation of the model. Further, by exploiting those occurrences, reasoning in lifted representations has no longer a complexity exponential in the number of objects represented by the model, here regions, but is limited to the number of objects with asymmetries only [11, 12]. More specifically, symmetries across objects of a models domain, i.e., objects over randvars of the same type, are exploited by means of performing calculations in inference only once for groups of similarly behaving objects, instead of performing the same calculations over and over again for all objects individually. The principle of lifting applies not only to logistics but also to many other areas like politics, healthcare, or finance – just to name a few.

DPRMs encode a temporal dimension and can be used in any *online scenario*, i.e., new knowledge is on the fly encoded to enable for continuous query answering without relearning the model. In existing research, it is assumed that lifted graphical models already contain symmetries, i.e., simply speaking, a model is setup so that all objects behave according to the same probability distribution. New knowledge is then incorporated in the model with new observations for each object. Observations are encoded within the model as realisations of randvars, resulting in a split off from a symmetrical consideration, called *grounding*. Of course, if the same observation is made for multiple objects, those objects are split off together and continue to be treated as a group. Over time, models dissolve into groups of symmetrically behaving objects, i.e., symmetries are implicitly exploited. Note that in the worst case, the models are split in such way that all objects are treated individually, i.e., no symmetries are available in the model so that lifted inference can no longer be applied and all its advantages disappear.

To the best of our knowledge, existing research has not yet fo-

cused on constructing symmetries in advance instead of deriving symmetries implicitly. Constructing symmetries in advance has benefits in application and results from the characteristics of real-world applications:

- (i) Certain information about objects of the model may not be available at runtime but only become known downstream. In such cases it is beneficial to infer information according to the open-world assumption from the behaviour of other object which tend to behave similar, i.e., applying an intrinsic default. On the one hand wrong information can be introduced in the model, but on the other it is likely that objects continue to behave the same as per other objects.
- (ii) Symmetrical objects can behave the same in the long term, but may deviate for shorter periods of time. Even already small deviations lead to groundings in the model, which, if prevented, introduce a small error in the model, but which also is negligible in the long term.

In both cases a model grounds, which must be prevented in order to keep reasoning in polynomial time. We construct groups of objects with similar behaviour, which we denote as *symmetry clusters*, through a symbolisation scheme to approximate temporal symmetries, i.e., objects that tend to behave the same over time. Using the symmetry clusters, it is possible to selectively prevent groundings, which helps to retain a lifted representation.

This work contributes with a summary on approximating model symmetries in DPRMs based on multivariate ordinal pattern symbolisation and clustering to obtain groups of objects with approximately similar behaviour. Behaviour is derived from the realisations of randvars, which generate either a univariate or multivariate time series depending on the number of interdependent randvars in the model. In the first original conference paper [1], we present multivariate ordinal pattern symbolisation for symmetry approximation (MOP₄SA) for the univariate case and introduce symmetry approximation for preventing groundings (SA₄PG) as an algorithm to prevent groundings in inference a priori using the learned entity symmetry clusters. In the second original conference paper [2], we extend MOP₄SA to the multivariate case and motivate the determination of related structural changes and periodicities in symmetry structures. Further, this work contributes with an extension of original works in [1] and [2] by

- a **comprehensive review** of MOP₄SA and SA₄PG with additional applications from dry-bulk shipping,
- a **complement** of the existing theoretical and experimental investigations of MOP₄SA and SA₄PG in the variation of its parameters, i.e., we fill in the gaps by investigating different orders and delays for ordinal patterns while also examining different thresholds in the symmetry approximation with respect to the introduced error in inference, and
- a **new approach** named MOP₄SCD to detect changes in symmetry structures based on a models similarity graph, an intermediate step of MOP₄SA, to re-learn symmetries on demand.

Together MOP₄SA, SA₄PG and multivariate ordinal pattern symbolisation for symmetry change detection (MOP₄SCD) combine into a rich toolset to identify model symmetries as part of the model

construction process, use those symmetries to maintain a lifted representation by preventing groundings a priori and detect changes in model symmetries after the model construction process.

This paper is structured as follows. After presenting preliminaries on DPRMs in Section 2, we continue in Section 3 with an review on how to retain lifted solutions through approximation and its relation to time series analysis and common approaches to determine and approximate similarities in that domain. In the following, we recapitulate MOP₄SA, an approach that encodes entity behaviour through ordinal pattern symbolisation in Section 4.1 and summarise approximating entity symmetries based on the symbolisation in Section 4.2. In Section 5 we outline SA₄PG and elaborate how to prevent groundings in DPRMs a priori with the help of entity symmetry clusters. In Section 6 we evaluate both MOP₄SA and SA₄PG in a shipping application and provide a detailed discussion on its various parameters and its effect on the accuracy in inference. In Section 7 we introduce a new approach to detect changes in symmetry structures to uncover points in time at which relearning of symmetry clusters becomes beneficial. In Section 9 we conclude with future work.

2 Background

In the following, we recapitulate DPRMs [5] in context of an example in logistics, specifically dry-bulk shipping. Dry-bulk shipping is one of the most important forms of transportation as part of the global supply chain [7, 3]. Especially the last year 2020, which was marked by the coronavirus pandemic, showed the importance of good supply chain management. The global supply chain was heavily affected as a result of required lock-downs all over the world, which has led to disruptions and significant delays in delivery. An important sub-challenge in supply chain management is to avoid congestion in regions/zones in which cargo is loaded, i.e., making sure that vessels arrive when those regions are not blocked by too many other vessels being anchored up in same. Congestion avoidance has always been an important topic in research [8]-[10]. As follows, we setup a DPRM to infer idle times related to global supply for commodities. As such, we formally define DPRMs and elaborate on sparse representations and more efficient inference by exploiting symmetries to enable for faster decision making.

2.1 A Formal Model on Congestion in Shipping

We setup a simplified DPRM to model congestion resulting in idle times in different regions/zones across the globe. To infer idle times in certain zones, we use freight rates, a fee per ton, which is paid for the transportation of cargoes and differs across zones, as a driver for operators to plan their vessel movements. E.g., an increase in idle time in a zone can be caused by a high freight rate in that same zone, resulting in an higher interest for sending vessels due to the potential to gain higher profits due to high freight rates. Of course, even though freight rates might be higher, not every vessel operator will be able to conclude business in zones which are over-crowded or have higher waiting times increasing costs for lay time. Hence, to describe the interaction between waiting times and freight rates, the

idle condition and freight rates in a zone can be represented by one randvar. Freight rates itself are driven by the supply of commodities in zones represented by another randvar. Since idle conditions, freight rates and supply can be similar in multiple zones, we can develop a much smaller model and combine all randvars into one and parameterise these with a logvar to represent the set of all zones respectively. In this example one zone from the set of all zones is referred to as an object or entity, which we use interchangeably moving forward. Such a parameterised random variable is referred to as PRV for short.

Definition 2.1 (PRV) Let \mathbf{R} be a set of randvar names, \mathbf{L} a set of logvar names, Φ a set of factor names, and \mathbf{D} a set of entities. All sets are finite. Each logvar L has a domain $\mathcal{D}(L) \subseteq \mathbf{D}$. A constraint is a tuple $(\mathcal{X}, C_{\mathcal{X}})$ of a sequence of logvars $\mathcal{X} = (X_1, \dots, X_n)$ and a set $C_{\mathcal{X}} \subseteq \times_{i=1}^n \mathcal{D}(X_i)$. A PRV $A(L_1, \dots, L_n), n \geq 0$ is a construct of a randvar name $A \in \mathbf{R}$ combined with logvars $L_1, \dots, L_n \in \mathbf{L}$. Then, the term $\mathcal{R}(A)$ denotes the (range) values of a PRV A . Further, the term $lv(P)$ refers to the logvars and $rv(P)$ to the randvars in some element P . The term $gr(P|_C)$ denotes the set of instances of P with all logvars in P grounded w.r.t. constraint C .

The idea behind PRVs is to enable for combining objects with similar behaviour in a single randvar to come up with a sparse representation, introducing a technique called *lifting*. To model the interaction between idle times, freight rates and supply in zones across the globe, we use randvars *Idle*, *Supply* and *Rate* parameterised with a logvar Z representing zones, building PRVs $Idle(Z)$, $Supply(Z)$ and $Rate(Z)$. The domain of Z is $\{z_0, z_1, \dots, z_n\}$ and range values for all PRVs are $\{high, medium, low\}^2$. A constraint $C = (Z, \{z_1, z_2\})$ for Z allows to restrict Z to a subset of its domain, such as here to z_1 and z_2 . Using this constraint, the expression $gr(Idle(Z)|_C)$ evaluates to $\{Idle(z_1), Idle(z_2)\}$. To represent independent relations, PRVs are linked by a parametric factor (parfactor) to compactly encode the full joint distribution of the DPRM.

Definition 2.2 (Parfactor) We denote a parfactor g by $\phi(\mathcal{A})|_C$ with $\mathcal{A} = (A^1, \dots, A^n)$ a sequence of PRVs, $\phi : \times_{i=1}^n \mathcal{R}(A^i) \mapsto \mathbb{R}^+$ a function with name $\phi \in \Phi$, and C a constraint on the logvars of \mathcal{A} . A PRV A or logvar L under constraint C is given by $A|_C$ or $L|_C$, respectively. An instance is a grounding of P , substituting the logvars in P with a set of entities from the constraints in P . A parameterized model PRM G is a set of parfactors $\{g^i\}_{i=1}^n$, representing the full joint distribution $P_G = \frac{1}{Z} \prod_{f \in gr(G)} f$, where Z is a normalizing constant.

All PRVs are dependent on each other and therefore are combined through one parfactor

$$g^1 = \phi^1(Idle(Z), Rate(Z), Supply(Z)), \quad (1)$$

which denotes their joint probability distribution. We omit the concrete mappings of potentials to range values of ϕ^1 . To encode temporal behaviour, DPRMs follow the same idea as dynamic Bayesian networks (DBNs) with an initial model and a temporal copy pattern to describe model changes over time. DPRMs model a stationary process, i.e., changes from one time step to the next follow the same distribution.

²for sake of simplicity we only consider three range values here

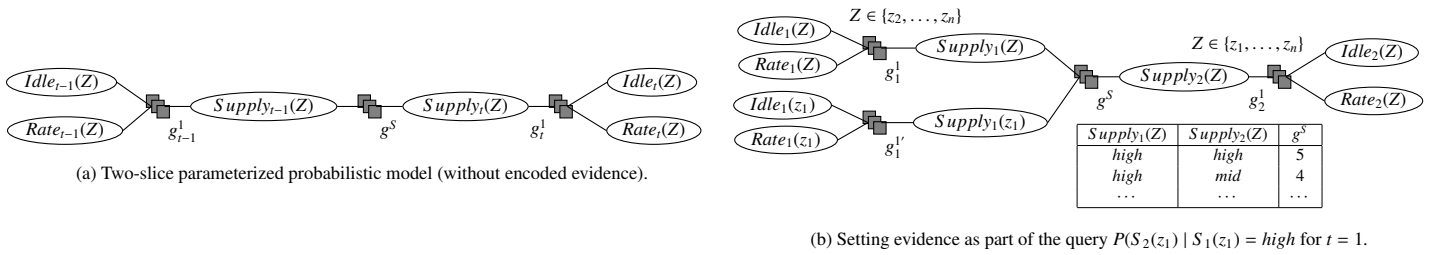


Figure 1: Graphical representation of a slice of a dynamic probabilistic graphical model illustration how to encode evidence.

Definition 2.3 (DPRM) A DPRM is a pair of PRMs (G_0, G_{\rightarrow}) where G_0 is a PRM representing the first time step and G_{\rightarrow} is a two-slice temporal parameterized model representing A_{t-1} and A_t where A_{π} is a set of PRVs from time slice π . An inter-slice parfactor $\phi(\mathcal{A})_C$ has arguments \mathcal{A} under constraint C containing PRVs from both A_{t-1} and A_t , encoding transitioning from time step $t - 1$ to t . A DPRM (G_0, G_{\rightarrow}) represents the full joint distribution $P_{(G_0, G_{\rightarrow}), T}$ by unrolling the DPRM for T time steps, forming a PRM as defined above.

Figure 1a shows the final DPRM. Variable nodes (ellipses) correspond to PRVs, factor nodes (boxes) to parfactors. Edges between factor and variable nodes denote relations between PRVs, encoded in parfactors. The parfactor g^S denotes a so-called inter-slice parfactor that separates the past from the present. The submodel on the left and on the right of this inter-slice parfactor are duplicates of each other, with the one on the left referring to time step $t - 1$ and the one on the right to time step t . Parfactors reference time-indexed PRVs, namely $Idle_t(Z)$, $Rate_t(Z)$ and $Supply_t(Z)$.

2.2 Query Answering under Evidence

Given a DPRM, one can ask queries for probability distributions or the probability of an event given evidence.

Definition 2.4 (Queries) Given a DPRM (G_0, G_{\rightarrow}) , a ground PRV Q_t , and evidence $E_{0:t} = \{\{E_{s,i} = e_{s,i}\}_{i=1}^n\}_{s=0}^t$ (set of events for time steps 0 to t), the term $P(Q_{\pi} | E_{0:t}, \pi \in \{0, \dots, T\}, t \leq T$, denotes a query w.r.t. $P_{(G_0, G_{\rightarrow}), T}$.

In context of the shipping application, an example query for time step $t = 2$, such as $P(Supply_2(z_1) | Supply_1(z_1) = \text{high})$, which asks for the probability distribution of supply at time step $t = 2$ in a certain zone z_1 , given that in the previous time step $t = 1$ the supply was high, contains an observation $Supply_1(z_1) = \text{high}$ as evidence. Sets of parfactors encode evidence, one parfactor for each subset of evidence concerning one PRV with the same observation.

Definition 2.5 (Encoding Evidence) A parfactor $g_e = \phi_e(E(X))_{C_e}$ encodes evidence for a set of events $\{E(x_i) = o\}_{i=1}^n$ of a PRV $E(X)$. The function ϕ_e maps the value o to 1 and the remaining range values of $E(X)$ to 0. Constraint C_e encodes the observed groundings x_i of $E(X)$, i.e., $C_e = (X, \{x_i\}_{i=1}^n)$.

Figure 1b depicts how evidence for $t = 1$, i.e., to the left of the interslice parfactor g^S , is set within the lifted model.

Evidence is encoded in parfactors g_1^l by duplicating the original parfactor g_1^l and using g_1^l to encode evidence and g_1^l to represent

all sets of entities that are still considered indistinguishable. Each parfactor represents a different set of entities bounded by the use of constraints, i.e., limiting the domain for the evidence parfactor g_1^l to $\{z_1\}$ and the domain for the original parfactor g_1^l to $\mathcal{D}(Z) \setminus \{z_1\}$. The parfactor that encodes evidence is adjusted such that all range value combinations in the parfactors distribution ϕ for $Supply_1(z_1) \neq \text{high}$ are dropped. Groundings in one time step are transferred to next time steps, i.e., also apply to further time steps, which we discuss as follows.

2.3 The Problem of Model Splits in Lifted Variable Elimination for Inference

As shown in Fig. 1b, evidence leads to groundings in the lifted model. Those model splits are carried over in message passing over time when performing query answering, i.e., in inference. Answering queries, e.g., asking for the probability of an event, results in joining dependent PRVs, or more specifically, joining those parfactors with overlapping PRVs. Figure 1b shows a sample of the probability distribution for the interslice parfactor g^S which separates time steps $t - 1$ and t . Answering the query $P(Supply_2(z_1) | Supply_1(z_1) = \text{high})$ as per the example in Section 2.1 to obtain the probability distribution over supply in time step $t = 2$, requires to multiply parfactors g_1^l with the interslice parfactor g^S and the parfactor g_2^l . However, since in time step $t = 1$ a grounding for z_1 exists, the grounding is carried over to the following time step $t = 2$ as the PRV $Supply_1(z_1)$ is connected to the following time step via the interslice parfactor g^S . Therefore, evidence, here $Supply_1(z_1) = \text{high}$, is also set within g^S dropping all range values for $Supply_1(z_1) \neq \text{high}$. This step is necessary to obtain an exact result in inference. For any other queries, this evidence is carried over to all future time steps, accordingly. Thus, under evidence a model $G_t = \{g_t^i\}_{i=1}^n$ at time step t is split w.r.t. its parfactors such that its structure remains

$$G_t = \{g_t^{i,1}, \dots, g_t^{i,k}\}_{i=1}^n \quad (2)$$

with $k \in \mathbb{N}^+$. Every parfactor g_t^i can have up to $k \in \mathbb{N}^+$ splits $g_t^{i,j} = \phi_t^{i,j}(\mathcal{A}^i)_{C^{i,j}}$, where $1 \leq j \leq k$ and \mathcal{A}^i is a sequence of the same PRVs but with different constraint $C^{i,j}$ and varying functions $\phi_t^{i,j}$ due to evidence. Note that moving forward we use the terms parfactor split or parfactor group interchangeably.

In our example, the model is only splitted with regards to evidence for the entity z_1 . All other entities are still considered to be indistinguishable, i.e., lifted variable elimination (LVE) can still exploit symmetries for those instances. To do so, lifted query answering is done by eliminating PRVs, which are not part of the

query, by so called *lifted summing out*. Basically, variable elimination is computed for one instance and exponentiated to the number of isomorphic instances represented. In [5], the author introduce the lifted dynamic junction tree (LDJT) algorithm for query answering on DPRMs, which uses LVE [6, 13] as a subroutine during its calculations. For a full specification of LDJT, we recommend to read on in [5]. In the worst case a model is fully grounded, i.e., a model as defined in Eq. (12) contains

$$k = \prod_{L \in \text{lv}(\mathcal{A})} |L| \quad (3)$$

splits for every parfactor $g_l^i = \phi_l^i(\mathcal{A})_{C^i}$ such that each object $l \in L$ is in its own parfactor split. The problem of model splits, i.e., groundings, can generally be traced back to two aspects. Groundings arise from

- (a) partial evidence or unknown evidence, i.e., certain information about objects of the model may not be available at runtime and either never or only become known downstream, which we denote as *unknown inequality*, or
- (b) from different observations for two or more objects, i.e., objects show different behaviour requiring to consider those individually moving forward, which we denote as *known inequality*.

Once the model is split, those splits are carried forward over time, potentially leading to a fully ground model. By doing so, the model remains exact as new knowledge (in form of observations) is incorporated into the model in all details. Over time, however, distinguishable entities might align and can be considered as one again (in case of known inequality) or entities might have ever since behaved similarly without knowing due to less frequent evidence (in case of unknown inequality). To retain a lifted representation the field of approximate inference, i.e., approximating symmetries, has emerged in research.

3 Related Work on Retaining Lifted Solutions Through Approximation and the Connection to Time Series Analysis

Lifted inference approaches suffer under the dynamics of the real world, mostly due to asymmetric or partial evidence. Handling asymmetries is one of the major challenges in lifted inference and crucial for its effectiveness [14, 15]. To address that problem, approximating symmetries has emerged in related research that we discuss in the following.

3.1 Approximate Lifted Models

For static (non-temporal) models, in [16] the author propose to approximate model symmetries as part of the lifted network construction process. They perform Lifted Belief Propagation (LBP) [17], which constructs a lifted network, and apply Belief Propagation (BP) to it. The lifted network is constructed by simulating message passing and identifying nodes sending the same message. To approximate the lifted network, message passing is stopped at

an earlier iteration to obtain an approximate instance. In [18], the author also approximate symmetries using LBP, but propose piecewise learning [19] of the lifted network. That means that the entire model is divided into smaller parts which are trained independently and then combined afterwards. In this way, evidence only influences the factors in each part, yielding a more liftable model. Besides approaches using LBP, in [20] the author propose evidence-based clustering to determine similar groundings in an Markov Logic Network (MLN). They measure the similarity between groundings and replace all similar groundings with their cluster centre to obtain a domain-reduced approximation. Since the model becomes smaller, also inference in the approximated lifted MLN is also much faster. In [15], the author propose so-called over-symmetric evidence approximation by performing low-rank boolean matrix factorisation (BMF) [21] on MLNs. They show, that for evidence with high boolean rank, a low-rank approximate BMF can be found. Simply put, finding a low-rank BMF corresponds to removing noise and redundant information from the data, yielding a more compact representation, which is more efficient as more symmetries are preserved. As with any existing approach to symmetry approximation, inference is performed on the symmetrised model, ignoring the introduction of potentially spurious marginals in the model. In [12], the author propose a general framework that provides improved probability estimates for an approximate model. Here, a new proposal distribution is computed using the Metropolis-Hasting algorithm [22, 23] on the symmetrised model to improve the distribution of the approximate model. Their approach can be combined with any existing approaches to approximate model symmetries.

Still, most of the existing research is based on static models and requires to get evidence in advance. However, the problem of asymmetric evidence is particularly noticeable in temporal models, and even more so in an *online setting*, since performing the symmetry approximation as part of the lifted network construction process is not feasible [24]. That means that it is necessary to construct a lifted temporal model once and to encode evidence as it comes in over time. Continuous relearning, i.e., reconstructing the lifted temporal model before performing query answering, is too costly. For temporal models in [25], the author propose to create a new lifted representation by merging groundings introduced over time. They perform clustering to group sub-models and perform statistical significance checks to test if groups can be merged.

In comparison to that and to the best of our knowledge, no-one has focused on preventing groundings before they even occur. To this end, we propose to learn entity behaviour in time and cluster entities that behave approximately similar in the long run and use them to accept or reject incoming evidence to prevent the model from grounding. Clustering entity behaviour requires approaches which find symmetries in entity behaviour, i.e., clustering entities which tend to behave the same according to observations made for them. As observations collected over time result in a time series our problem comes down to identify symmetries across time series.

3.2 From DPRMs to Time Series

In a DPRM, (real-valued) random variables observed over time are considered as time series. Let Ω be a set containing all possible states of the dynamical system, also called state space. Events are

taken from a σ -algebra \mathcal{A} on Ω . Then (Ω, \mathcal{A}) is a measurable space. A sequence of random variables, all defined on the same probability space $(\Omega, \mathcal{A}, \mu)$, is called a *stochastic process*. For real-valued random variables, a stochastic process is a function

$$X : \Omega \times \mathbb{N} \rightarrow \mathbb{R}, \quad (4)$$

where $X(\omega, t) := X_t(\omega)$ depending on both, coincidence and time. Note that in the most simple case Ω matches with \mathbb{R} and X with the identity map. Then the observations are directly related to iterates of some ω , i.e., there is no latency, and the X itself is redundant. Over time, the individual variables $X_t(\omega)$ of this stochastic process are observed, so-called realisations. The sequence of realisations is called *time series*. With the formalism from above and fixing of some $\omega \in \Omega$, a time series is given by

$$(X_1(\omega), X_2(\omega), X_3(\omega), \dots) = (x_t)_{t \in \mathbb{N}}. \quad (5)$$

In the case $x_t \in \mathbb{R}$ the time series is called univariate, while in the case $x_t \in \mathbb{R}^m$ it is called multivariate. Note that for stochastic processes we use the capitalisation $(X(t))_{t \in \mathbb{N}}$, while for observations, i.e., paths or time series, we use the small notation $(x(t))_{t \in \mathbb{N}}$. In summary, evidence in a DPRM encoding stochastic processes $(X(t))_{t \in \mathbb{N}}$ forms a time series $(x_t)_{t \in \mathbb{N}}$ that is the subject of further consideration.

3.3 Symmetry Approximation in Time Series

In time series analysis, the notion of similarity, known as symmetry in DPRMs, has often been discussed in the literature [26]-[28]. In general, approaches for finding similarities in a set of time series are either (a) value-based, or (b) symbol-based. *Value-based* approaches compare the observed values of time series. By comparing the value of each point $x_t, t = 1, \dots, T$ in a time series X with the values of each other point $y_{t'}, t' = 1, \dots, T'$ in another time series Y (warping), they are able to include shifts and frequencies. Popular algorithms such as dynamic time warping (DTW) [29] or matrix profile [30] are discussed, e.g., in [28]. As DPRMs can encode interdependencies between multiple variables, respective multivariate procedures should be used to assess similarities. The first dependent multivariate dynamic time warping (DMDTW) approach is reported by [31], in which the authors treat a multivariate time series with all its m interdependencies as a whole. The flexibility of warping in value-based approaches leads to a high computational effort and is therefore unusable for large amounts of data. Although there are several extensions to improve runtime [32] by limiting the warping path or reducing the number of data points, e.g., FastDTW [32] or PrunedDTW [33], the use of dimensionality reduction is inevitable in context of DPRMs. For dimensionality reduction, *symbol-based* approaches encode the time series observations as sequences of symbolic abstractions that match with the shape or structure of the time series. Since DPRMs encode discrete values, depending on the degree for discretisation, symbol-based approaches are preferred as they allow for discretisation directly. As far as research is concerned, there are two general ways of symbolisation. On the one hand, *classical symbolisation* partitions the data range according to specified mapping rules in order to encode a numerical time series into a sequence of discrete symbols. A corresponding and well-know algorithm is Symbolic Aggregate AppRoXimation (SAX)

introduced by [34]. On the other hand, as introduced by Bandt and Pompe [35] *ordinal pattern symbolisation* encodes the total order between two or more neighbours ($x < y$ or $x > y$) into so-called ordinal symbols ((0, 1) or (1, 0)). In [36], the author extend univariate ordinal patterns to the multivariate case, taking into account not only the dependencies of neighbouring values over time, but also the dependencies between spatial variables in a time series.

Specifically here, an ordinal approach has notable advantages in application: (i) The method is conceptionally simple, (ii) the ordinal approach supports robust and fast implementations [37, 38], and (iii) compared to classical symbolisation approaches such as SAX, it allows an easier estimation of a good symbolisation scheme [39, 40]. In the following, we introduce ordinal pattern symbolisation to classify similar entity behaviour.

4 Multivariate Ordinal Pattern for Symmetry Approximation (MOP₄SA)

In this section we recapitulate MOP₄SA, an approach for the approximation of symmetries over entities in the lifted model. MOP₄SA consists of two main steps, which is (a) encoding entity model behaviour through an *ordinal pattern symbolisation* approach, followed by (b) clustering entities with a similar symbolisation scheme to determine groups of entities with *approximately similar behaviour*. We have introduced MOP₄SA in [1] for the univariate case and extended same in [2] to the multivariate case.

4.1 Encoding Entity Behaviour through Ordinal Pattern Symbolisation

As mentioned in Section 3.3, approximating entity behaviour corresponds to finding symmetries in time series.

4.1.1 Gathering Evidence

To find symmetries in (multivariate) time series, we use evidence which encode model entity behaviour w.r.t. a certain context, i.e., w.r.t. a parfactor. In particular, this means: Every time-index PRV $P_t(X)$ represents multiple entities x_0, \dots, x_n of the same type at a specific point in time t . That is, for a PRV $Supply_t(Z)$, zones z_0, \dots, z_n are represented by a logvar Z with domain $\mathcal{D}(Z)$ and size $|\mathcal{D}(Z)|$. Note that a PRV can be parameterised with more than one logvar, but for the sake of simplicity we introduce our approach using PRVs with only one logvar throughout this paper. Symmetry detection for m -logvar PRVs works similarly to one-logvar PRVs, with the difference, that in symmetry detection, entity pairs, i.e., m -tuples, are used. As an example, for any 2-logvar PRV $P_t(X, Y)$, an entity pair is a 2-tuple (x_1, y_1) with $x_1 \in \mathcal{D}(X)$ and $y_1 \in \mathcal{D}(Y)$.

A DPRM, as introduced in Section 2.1, encodes temporal data by unrolling a DPRM while observing evidence for the models PRVs, e.g., the PRV $Supply_t(Z)$ encodes supply at time t in various zones Z on the globe. In addition, a DPRM exploits (conditional) interdependencies between randvars by encoding interdependencies in parfactors. As such, parfactors describe interdependent data through its linked PRVs, e.g., the correlation between supply $Supply_t(Z)$, idle times $Idle_t(Z)$ and freight rates $Rate_t(Z)$ within a common zone

Z encoded by the parfactor g_t^1 . For each entity $z_i \in \mathcal{D}(Z)$ from the PRVs $P = \{Supply_t(Z), Idle_t(Z) \text{ and } Rate_t(Z)\}$ observations are made over time, i.e., a time series $((x_t^i)_{i=1}^m)_{t=1}^T$ with $x_t^i \in \mathcal{R}(P^i)$ is generated. In this work, the time series is to be assumed multivariate, containing interdependent variables, i.e., $m > 1$. Note that in [1] we consider the case $m = 1$. Having $|\mathcal{D}(Z)|$ entities in Z, we consider $|\mathcal{D}(Z)|$ samples of multivariate time series

$$\mathcal{X} = (((x_t^i)_{i=1}^m)_{t=1}^T)_{j=1}^{|\mathcal{D}(Z)|} \in \mathbb{R}^{m \times T \times |\mathcal{D}(Z)|}, \quad (6)$$

e.g., for $m = 3$ with observations $(x_t^1, x_t^2, x_t^3) = (Supply_t(z_j), Idle_t(z_j), Rate_t(z_j))$ for every $z_j \in \mathcal{D}(Z)$ in time $t \in \{1, \dots, T\}$. As such, a multivariate time series is defined for several PRVs linked in a parfactor, while a univariate time series is defined for a single PRV. Identification of symmetrical entity behaviour is done on a sets of (multivariate) time series, i.e., across different (multivariate) time series that are observed for every entity individually.

4.1.2 Multivariate Ordinal Pattern (MOP) Symbolisation

To encode the behaviour of a time series, we use ordinal pattern symbolisation based on works from Bandt and Pompe [35]. For this, let $X_t \in \mathbb{R}^{m \times T}$ be a (multivariate) time series and $X_r \in \mathbb{R}^{m \times T \times n}$ be the reference database of $n \in \mathbb{N}$ (multivariate) time series. In case of $m = 1$, the time series is univariate. For a better understanding, we start with univariate ordinal patterns that encode the up and downs in a time series by the total order between two or more neighbours. The encoding gives a good abstraction, an approximation, of the overall behaviour or generating process. Univariate ordinal patterns are formally defined as follows.

Definition 4.1 (Univariate Ordinal Pattern) A vector $(x_1, \dots, x_d) \in \mathbb{R}^d$ has ordinal pattern $(r_1, \dots, r_d) \in \mathbb{N}^d$ of order $d \in \mathbb{N}$ if $x_{r_1} \geq \dots \geq x_{r_d}$ and $r_{l-1} > r_l$ in the case $x_{r_{l-1}} = x_{r_l}$.

Figure 2 shows all possible ordinal patterns of order $d = 3$ of a vector $(x_1, x_2, x_3) \in \mathbb{R}^3$.

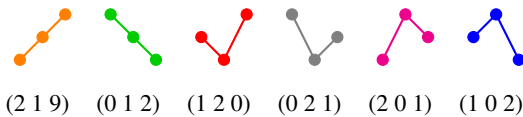


Figure 2: All $d!$ possible univariate ordinal patterns of order $d = 3$.

For a multivariate time series $((x_t^i)_{i=1}^m)_{t=1}^T$, each variable x^i for $i \in 1, \dots, m$ depends not only on its past values but also has some dependency on other variables. To establish a total order between two time points $(x_t^i)_{i=1}^m$ and $(x_{t+1}^i)_{i=1}^m$ with m variables is only possible if $x_t^i > x_{t+1}^i$ or $x_t^i < x_{t+1}^i$ for all $i \in 1, \dots, m$. Therefore, there is no trivial generalisation to the multivariate case. An intuitive idea, based on some theoretical discussion in [41, 42] and introduced in [36], is to store univariate ordinal patterns of all variables at a time point t together into a symbol.

Definition 4.2 (Multivariate Ordinal Pattern) A matrix $(x_1, \dots, x_d) \in \mathbb{R}^{m \times d}$ has multivariate ordinal pattern (MOP) of order $d \in \mathbb{N}$

$$\begin{pmatrix} r_{11} & \dots & r_{1d} \\ \vdots & \ddots & \vdots \\ r_{m1} & \dots & r_{md} \end{pmatrix} \in \mathbb{N}^{m \times d} \quad (7)$$

if $x_{r_{i1}} \geq \dots \geq x_{r_{id}}$ for all $i = 1, \dots, m$ and $r_{i,l-1} > r_{i,l}$ in the case $x_{r_{i,l-1}} = x_{r_{i,l}}$.

For $m = 1$ the multivariate case matches with the univariate case in Definition 4.1. Figure 3 shows all $(d!)^m$ possible multivariate ordinal patterns (MOPs) of order $d = 3$ and number of variables $m = 2$.

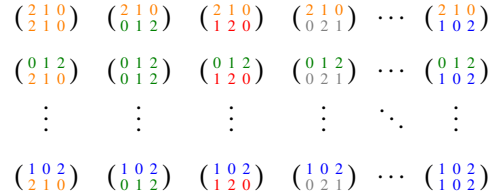


Figure 3: All $(d!)^m$ possible multivariate ordinal patterns of order $d = 3$ with $m = 2$ variables.

The number of possible MOPs $d!$ increases exponentially with the number of variables m , i.e., $(d!)^m$. Therefore, if d and m are too large, depending on the application, each pattern occurs only rarely or some not at all, resulting in a uniform distribution of ordinal patterns [36]. This has the consequence that subsequent learning procedures can fail. Nevertheless, for a small order d and sufficiently large T the use of MOPs can lead to higher accuracy in learning tasks, e.g., classification [36] because they incorporate interdependence of the spatial variables in the multivariate time series.

To symbolise a multivariate time series $X_t \in \mathbb{R}^{m \times T}$ each pattern is identified with exactly one of the ordinal pattern symbols $o = 1, 2, \dots, d!$, before each point $t \in \{d, \dots, T\}$ is assigned its ordinal pattern symbol of order $d \ll T$. The order d is chosen much smaller than the length T of the time series to look at small windows in a time series and their distributions of up and down movements. To assess long-term trends, delayed behaviour is of interest, showing various details of the structure of the time series. The time delay $\tau \in \mathbb{N}_{>0}$ is the delay between successive points in the symbol sequences.

4.1.3 MOP Symbolisation with Data Range Dependence

We assume that for each time step $t = \tau(d - 1) + 1, \dots, T$ of a multivariate time series $((x_t^i)_{i=1}^m)_{t=1}^T \in \mathcal{X}$, MOP is determined as described in Section 4.1.2. Ordinal patterns are well suited to characterise an overall behaviour of time series, in particular their application independent of the data range. In some applications, however, the dependence on the data range can be also relevant, i.e., time series can be similar in terms of their ordinals patterns, but differ considering their y-intercept. In other words, transforming a sequence

$$x = (x_t^i)_{a \leq t \leq b} \quad (8)$$

as $y = x + c$, where $c \in \mathbb{R}$ is a constant, should change y 's similarity to other sequences, although the shape is the same. To address the

dependence on the data range, we use the arithmetic mean

$$\bar{x}_t^{d,\tau} = \frac{1}{m} \sum_{i=1}^m \frac{1}{d} \sum_{k=1}^d x_{i,t-(k-1)\tau} \quad (9)$$

of the multivariate time series' values corresponding to the ordinal pattern, where $x_{i,t-(k-1)\tau}$ is min-max normalised, as an additional characteristic or feature of behaviour. If one of the variables changes its behaviour significantly along the intercept, the arithmetic mean uncovers this. There are still other features that can be relevant. For simplicity, we only determine ordinal patterns and their means for each parfactor g^1 with, e.g., PRVs ($Supply_t(Z)$, $Idle_t(Z)$, $Rate_t(Z)$), yielding a new data representation

$$\mathcal{X}' = \langle o, \bar{x} \rangle^{(T-(\tau(d-1))) \times |\mathcal{D}(Z)|)} \quad (10)$$

where $\langle o, \cdot \rangle_{ij} \in \mathcal{X}'$ represents the MOP and $\langle \cdot, \bar{x} \rangle_{ij} \in \mathcal{X}'$ represents the corresponding mean $\bar{x}_t^{d,\tau}$ for entity z_j at time step t . The order d and delay τ are passed in from the outside and might depend on, e.g., the frequency of the data, to capture the long-term behaviour.

4.2 Clustering Entities with Similar Symbolisation Scheme

After encoding the behaviour of the entities through ordinal pattern symbolisation, we identify similar entities using clustering. For this purpose, based on the derived symbolisation representation in Eq. (10), we create a similarity graph indicating the similarities based on a distance measure between entity pairs.

4.2.1 Creating a Similarity Graph

Entity similarity is measured per parfactor, i.e., per multivariate time series, separately. Therefore, multiple similarity graphs, more specifically one per parfactor, are constructed. A similarity graph for a parfactor g_t^1 connecting the PRVs $Supply_t(Z)$, $Idle_t(Z)$ and $Rate_t(Z)$ contains one node for each entity $z \in \mathcal{D}(Z)$ observed in form of multivariate time series. The edges of the similarity graph represent the similarity between two nodes, or more precisely, how closely related two entities of the model are. To measure similarity, we use the symbolic representation \mathcal{X}' , which contains tuples of multivariate ordinal numbers and mean values that describe the behaviour of an entity. The similarity of two entities z_i and z_j is given by counts w_{ij} of equal behaviours, i.e.,

$$w_{ij} = \sum_{t \leq T} \left[\langle o, \cdot \rangle_{it} = \langle o, \cdot \rangle_{jt} \wedge | \langle \cdot, \bar{x} \rangle_{it} - \langle \cdot, \bar{x} \rangle_{jt} | < \delta \right], \quad (11)$$

where $[x] = 1$ if x and, 0 otherwise. As an auxiliary structure, we use a square matrix $\mathcal{W} \in \mathbb{N}^{|\mathcal{D}(Z)| \times |\mathcal{D}(Z)|}$, where each $w_{ij} \in \mathcal{W}$ describes the similarity between entities z_i and z_j by simple counts of equal behaviour over time $t \in T$. Simply put, one counts the time steps t at which both multivariate time series of z_i and z_j have the same MOP and the absolute difference of the mean values of the corresponding MOPs is smaller than $\delta > 0$. Finally, as shown in Figure 4b the counts w_{ij} correspond to the weights of edges in the similarity graph \mathcal{W} , where zero indicates no similarity between two entities, while the larger the count, the more similar two entities are.

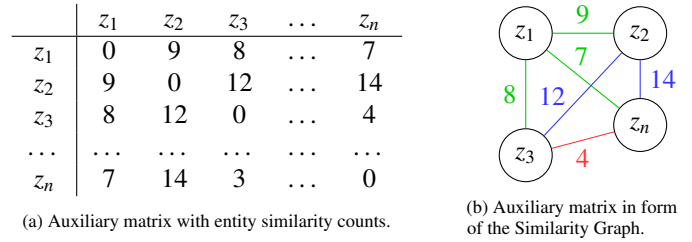


Figure 4: (a) Auxiliary matrix and (b) similarity graph.

Approximating symmetries based on the similarity graph leaves us with a classical clustering problem. That means, clustering entities into groups of entities showing *enough* similarities or leaving others independent if those are too different.

4.2.2 Derive Entity Similarity Clusters

Theoretically, any clustering algorithm can be applied on top of the similarity graph. Each weight in the similarity matrix, or each edge weight between an entity pair, denotes the similarity between two entities, i.e., the higher the count, the more similar the two entities are to each other. Since this contribution focuses on identifying symmetries in temporal environments, we leave the introduction of a specific clustering algorithm out here, and compare two different ones, specifically Spectral Clustering and Density-Based Spatial Clustering of Applications with Noise (DBSCAN), for the use in MOP₄SA as part of the evaluation in Section 6. After any clustering algorithm is run, we are left with n clusters of entities for each parfactor. Formally, a symmetry cluster is defined as follows.

Definition 4.3 (Symmetry Cluster) For a parfactor $g_i \in G_t$ with G_t being the PRM at timestep t and $g_i = \phi(\mathcal{A})_C$ containing a sequence of PRVs $\mathcal{A} = (A^1, \dots, A^n)$, a symmetry cluster S^i contains entities $l \in \mathcal{D}(L)$ concerning the domain $\mathcal{D}(L)$ of one of the logvars $L \in \mathcal{L}$ with $\mathcal{L} = \bigcup_{i=1}^n \text{lv}(A^i)$. Let the term $en(S)$ refer to the set of entities in any symmetry cluster S . Each parfactor $g_i \in G_t$ can contain up to m symmetry clusters $\mathcal{S}_{|g^i} = \{S^i\}_{i=0}^m$, such that $en(S^i) \cap en(S^j) = \emptyset$ for $i \neq j$ and $i, j \in \{1, \dots, m\}$. $|\top$ may be omitted in $\mathcal{S}_{|\top}$.

In the following section we propose how to utilise symmetry clusters to prevent any lifted model from unnecessary groundings.

5 Symmetry Approximation for Preventing Groundings (SA₄PG)

As described in Section 2.3, evidence leads to groundings in any lifted model. Further, those groundings are carried over in message passing when moving forward in time leading to a fully ground model in the worst case. As follows, we propose SA₄PG, which uses symmetry clusters as an outcome of MOP₄SA to counteract any unnecessary groundings, which occur due to evidence. Since symmetry clusters denote a sets of entities, for which entities in each group tend to behave the same, also observations for each entity individually within a cluster are expected to be similar. Regardless of our approach to prevent groundings, in DPRMs entities are

considered indistinguishable after the initial model setup. Under evidence entities split off from that indistinguishable consideration and are afterwards treated individually to allow for exact inference. Nevertheless, in case one observation was made for multiple entities, those all together split off and are considered individually, but still within the group of entities for which the that observation was made. Such groundings are encoded within the DPRM in parfactor groups as shown in Eq. (12). Symmetry clusters also denote parfactor groups, with the difference that those are determined as part of the model construction process in advance. Therefore, in the model construction process, i.e., before running inference, a model will be splitted according to the clusters into parfactor groups with each group containing only entities from the respective cluster. The only difference in creating parfactor groups without evidence is that no range values are set to zero, but get a different distribution representing the group the best. SA₄PG is based on the assumption, that symmetry clusters *stay valid* for a certain period of time after learning them, i.e., that entities within those clusters continue to behave similarity. More specifically and w.r.t. the two types of inequality (see Section 2.3), this means, that

- (a) in case of *unknown inequality*, we assume that any entity without an observation most likely continues to behave similar to the other entities within the same cluster for which observations are present, and
- (b) in case of *known inequality*, we assume that certain observations dominate one cluster and therefore will be applied for all entities within the cluster.

To make one example, lets assume a symmetry cluster contains entities z_1, z_2 and z_3 . Groundings occur whenever observations differ across entities in a symmetry cluster, e.g., grounding occurs, if (a) the observation ($Supply_1(z_1) = high, Idle_1(z_1) = high$) of entity z_1 differs from observations ($Supply_1(z_i) = low, Idle_1(z_i) = mid$) of entities z_i for $i = 2, 3$, or (b) observations are only made for a subset of the entities, i.e., for entities z_2 and z_3 , but not for entity z_1 . In both cases, the entities z_2 and z_3 would be split off from their initial symmetry group, and are henceforth treated individually in a non lifted fashion. In SA₄PG we prevent such model splits until a certain extend. Algorithm 1 shows an outline of the overall preventing groundings approach. Preventing groundings works by consuming evidence and queries from a stream and dismissing or inferring evidence within symmetry clusters until an entity has reached an violation threshold H . The threshold H refers to the number of times evidence was inferred or dismissed. To do not force entities to stick to their initial symmetry clusters, we relieve entities from their clusters once the threshold H is received. To keep track on the number of violations, i.e., how often evidence was inferred or dismissed, we introduce a violation map as a helper data structure to store that number.

Definition 5.1 (Violation Map) For a parfactor $g_i \in G_t$ with G_t being the PRM at timestep t and $g_i = \phi(\mathcal{A})_C$ containing a sequence of PRVs $\mathcal{A} = (A^1, \dots, A^n)$, a violation map $v_{|g_i} : \bigcup_{i=1}^n gr(A^i) \rightarrow 0$ is initialised with zero values for all entities in all PRVs \mathcal{A} in g_i . In case a PRV A^i is parameterised with more than one logvar, i.e., $m = |lv(A^i)|$ with $m > 1$, v contains m -tuples as entity pairs. A model contains up to n parfactors in G_t . The set of violation maps

is denoted by $V = \{v_{|g_i}\}_{i=0}^n$. Let $viol(P)$ refer to the violation count of some entity m -tuple in V .

SA₄PG continues by taking all evidence \mathbf{E}_t concerning a timestep $t = 0, 1, \dots, T$ from the evidence stream \mathcal{E} . To set evidence and to prevent groundings, for each observation $E_{s,i} \in \mathbf{E}_t$ with $\mathbf{E}_t = \{E_{s,i} = e_{s,i}\}_{i=1}^n$ so called *parfactor partitions* are identified. A parfactor partitions is a set of parfactor groups $g_t^{i,k}$ that are all affected by evidence $E_{s,i}(x_j)$ with $x_j \in \mathcal{D}(lv(E_{s,i}))$. A parfactor group is *affected*, if

- (a) the parfactor g_t^i itself links the PRV $E_{s,i}$ for which an observation was made,
- (b) and if the parfactor group g_t^i currently represents the distribution for the specific entity x_j for which the observation was made.

To make one example, observing $Supply_1(z_1) = high$, the evidence partition contains parfactor groups of the parfactors g_t^1 and g^S since the PRV is linked to both parfactors. Further, the parfactor partition is limited to only those i parfactor groups $g_t^{i,1}$ and $g_t^{i,S}$, which currently represent the distribution for the entity z_1 . A parfactor partition containing all those parfactor groups is defined as follows.

Definition 5.2 (Parfactor Partition) Every parfactor $g_t^i \in G_t$ can have up to $k \in \mathbb{N}^+$ splits such that

$$G_t = \{g_t^{i,1}, \dots, g_t^{i,k}\}_{i=1}^n \quad (12)$$

Each parfactor g_t^i contains a sequence of PRVs $\mathcal{A}_t = (A_t^1, \dots, A_t^n)$, which are afflicted with evidence $A_t^n(x_i) = a_{t,i}$ for any entity $x_i \in \mathcal{D}(X)$ with $X \in lv(\mathcal{A}_t)$ at timestep t leading to those splits. A parfactor partition P_t denotes a set of parfactors, which are affected by new evidence $E_t(x_i) = e_t$ with

$$P_t = \{g_t^{i,1}, \dots, g_t^{i,l}\}_{i=1}^n \quad (13)$$

and $l \leq k$ such that any parfactor group $g_t^{i,l} \in P_t$ contain the random var E_t , i.e., $E_t \in rv(g_t^{i,l})$ and $g_t^{i,l}$ is limited by constraints to at least the entity x_i for which the observation was made, i.e., $g_t^{i,l}|_{C_e}$ with $C_e = (X, \{x_i\}_{i=1}^n)$ and $x_i \in \{x_i\}_{i=1}^n$.

Considering all evidence \mathbf{E}_t for a time step t , different observations $E_{t,i} \in \mathbf{E}_t$ can result in the same parfactor partition (before those observations are encoded within the model). This holds true for all observation, which are made for the same PRV with entities being in the same parfactor group, e.g., two observations $Supply_1(z_i) = high$ and $Supply_1(z_j) = mid$ for which $\{z_i, z_j\} \in gr(g_t^{1,l})$ and $\{z_i, z_j\} \in gr(g_t^{S,l})$. All observations that entail the same parfactor partition are treated in SA₄PG as one and those observations are informally denoted as an *evidence cluster*.

Therefore, in SA₄PG evidence \mathbf{E}_t is rearranged in a sense such that \mathbf{E}_t contains multiple collections of observations, i.e.,

$$\mathbf{E}_t = \{\{E_{t,l} = e_{t,l}\}_{l=0}^m, \dots, \{E_{t,l} = e_{t,l}\}_{l=0}^m\}, \quad (14)$$

with each element $E_{t,l}$ originally being directly in \mathbf{E}_t and each subset $\{E_{t,l} = e_{t,l}\}_{l=0}^m$ concerning the same parfactor partition P_t . SA₄PG proceeds by processing each evidence cluster separately. Evidence of each evidence cluster is processed in a sense such that known inequalities and any uncertainty about inequality is counteracted. This is being done by identifying the *dominating observation* within

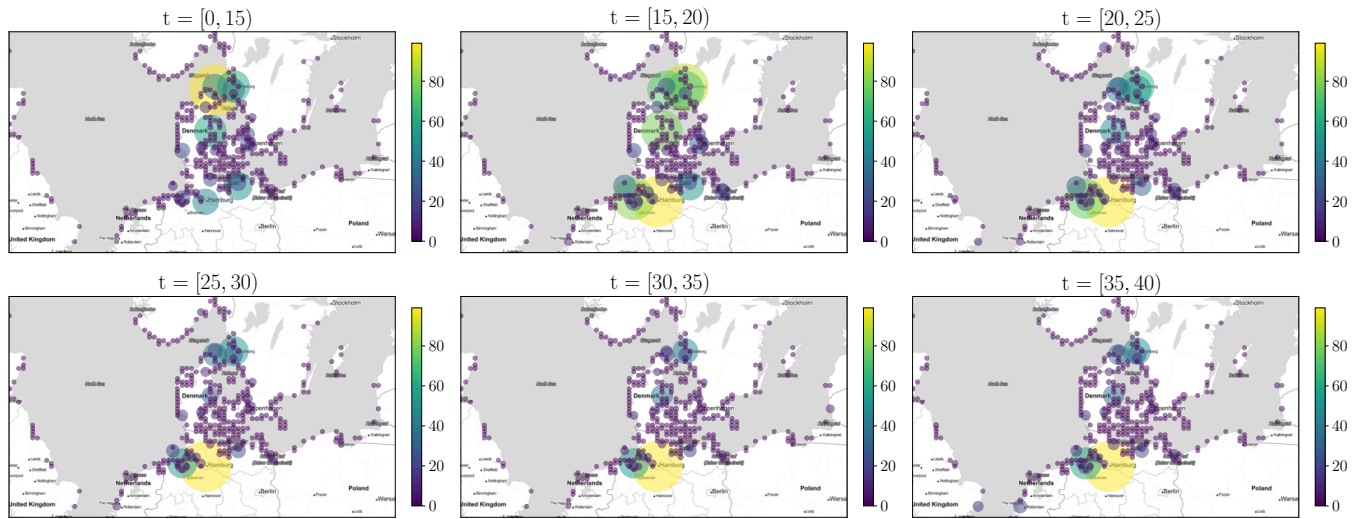


Figure 5: Pointmap showing the normalised average supply over time intervals $[t - 5, t)$ in the Baltic Sea region. Best viewed in colour.

each evidence cluster $\{E_{t,l} = e_{t,l}\}_{l=0}^m$ and apply that observation to all entities within the respective parfactor partition P_t . The dominating observation $\max(e_{t,l})$ is the observation that can be observed the most within the evidence cluster. Further, in case any other entities in the corresponding parfactor partition are unobserved, we also apply the dominating observation for those. For each entity for which evidence was inferred or dismissed, the violation counter is increased. In case the violation threshold H for an entity is already reached, evidence is no longer inferred or dismissed, but the entity is relieved from its symmetry cluster, i.e., split off from the parfactor group, and from then on considered individually.

In the following, we evaluate MOP₄SA and SA₄PG as part of a case study from the example shipping domain.

6 MOP₄SA and SA₄PG in Application

Since MOP₄SA consists of multiple steps, namely (a) encoding entity behaviour, (b) similarity counting, and (c) clustering, we evaluate each step separately before analysing the overall fitness in conjunction with SA₄PG, as introduced in Section 5.

6.1 The AIS Dataset

To setup a DPRM as shown in Figure 1a, we use historical vessel movements from 2020 based on automatic identification system (AIS) data² provided by the Danish Maritime Authority for the Baltic Sea. AIS data improves the safety and guidance of vessel traffic by exchanging navigational and other vessel data. It was introduced as a mandatory standard by the International Maritime Organisation (IMO) in 2000. Meanwhile, AIS data is used in many different applications in research, such as trade flow estimation, emission accounting, and vessel performance monitoring [43]. Pre-processing for retrieving variables *Supply* and *Idle* for 367 defined Zones can be found on GitHub³. Figure 5 gives an idea on how supply evolves over time t in the Baltic Sea region. Each point

illustrates the normalised cargo supply amount in tons. For sake of simplicity, we only plot supply independent of idle times here. We can see, that in the beginning of the year (for $0 < t < 20$) the supply in the northern regions, i.e. the need for resources, is higher, while for the rest of the year (for $20 < t < 40$) the supply slowly decreases and increases in the southern regions. The important part here is, that the supply for $20 < t < 40$ in the respective regions is more or less constant over a longer period of time.

Algorithm 1: Preventing Groundings

Input: DPRM (G_0, G_{\rightarrow}) , Evidence- \mathcal{E} and Querystream \mathcal{Q} , Order d , Delay τ , Symmetry Clusters C

for each parfactor $g_i \in G_0$ **do**

$v_{|g_i} \leftarrow$ init violation map // see Definition 5.1

for $t = 0, 1, \dots, T$ **do**

$\mathbf{E}_t \leftarrow$ get evidence from evidence stream \mathcal{E}

 Rearrange \mathbf{E}_t to create evidence clusters according to parfactor partition P_t // see Definition 5.2

for each evidence cluster $\{E_{t,l} = e_{t,l}\}_{l=0}^m \in \mathbf{E}_t$ **do**

$\max(e_{t,l}) \leftarrow$ get dominating observation

 // Align Evidence

for each observation in $E_{t,l}(x_i) \in \{E_{t,l} = e_{t,l}\}_{l=0}^m$ **do**

if $e_{t,l} \neq \max(e_{t,l})$ and $\text{viol}(x_i) < H$ **then**

 Dismiss observation $e_{t,l}$

$\text{viol}(x_i) \leftarrow \text{viol}(x_i) + 1$

 // Infer Evidence

for each unobserved entity x_j in P_t **do**

 Set $E_{t,l}(x_j) = \max(e_{t,l})$

 Answer queries Q_t from query stream \mathcal{Q}

The idea behind MOP₄SA is to identify periods of time with similar behaviour for multiple entities. That means in our application, identifying zones with similar supply (or more specifically in the multivariate context supply/idle times) over a period of time.

²<https://www.dma.dk/SikkerhedTilSoes/Sejladsinformation/AIS/>

³<https://github.com/FinkeNils/Processed-AIS-Data-Baltic-Sea-2020-v2>

Next, we look into clustering based on the similarity graph as an outcome of similarity counting after applying the symbolisation.

6.2 Multivariate Symbolisation Scheme for Temporal Similarity Clustering

According to the procedure as introduced in Section 4.1.3, we apply the symbolisation scheme on the multivariate supply/idle-time time series as encoded in the parfactor g_i^j and create one similarity graph as the basis for clustering. We compare different clustering algorithms as follows. Unfortunately, classical clustering methods do not achieve good results in high-dimensional spaces, like for DPRMs, which are specifically made to represent large domains. Problems that classical clustering approaches have is, that the smallest and largest distances in clustering differ only relatively slightly in a high-dimensional space [44]. For DPRMs, a similarity graph, representing the similarity of entities $z \in \mathcal{D}(Z)$, contains

$$\binom{|\mathcal{D}(Z)|}{2} \quad (15)$$

fully-connected nodes in the worst case, where Z is a logvar representing a set of entities whose entity pairs share similar behaviour for least one time step. Here, Eq. (15) also corresponds to the number of dimensions a clustering algorithm has to deal with.

6.2.1 An Informal Introduction to Clustering

Generally, clustering algorithms can be divided into the four groups (a) centroid-based clustering, (b) hierarchical clustering, (c) graph-based clustering, and (d) density-based clustering.

We already pointed out the problem that classical clustering algorithms suffer due to their distance measures, which do not work well in high dimensional spaces. Especially centroid-based clustering approaches, like the well-known k -means algorithm or Gaussian Mixture Models, suffer, as they expect to find spherical or ellipsoidal symmetry. More specifically, in centroid-based clustering the assumption is that the points assigned to a cluster are spherical around the cluster centre and therefore no good clusters can be found due to the relatively equal distances. In hierarchical clustering time and space complexity is especially bad since the graph is iteratively split into clusters. Graph-based clustering algorithms, like spectral clustering, is known as being especially robust for high-dimensional data due to performing dimensionality reduction before running clustering [45]. One disadvantage, which also applies to clustering algorithms above, is that the number of clusters need to be specified as a hyperparameter in advance. In contrast, in density-based clustering approaches, like DBSCAN, the number of clusters are determined automatically while also handling noise. DBSCAN is based on a high-density of points. That means, clusters are dense regions, which are identified by running with a sliding window over dense points, making DBSCAN cluster shape independent.

For these reasons, we will compare spectral clustering and DBSCAN as part of MOP₄SA as follows. We start by informally introducing Spectral Clustering and DBSCAN.

DBSCAN works by grouping together points with many nearby neighbours, denoting points lying outside those regions as noise.

In DBSCAN the two parameters ϵ and $minPoint$ need to be provided from the outside, which correspond to the terms *Density Reachability* and *Density Connectivity* respectively. The idea behind DBSCAN is to identify points, that are reachable from another if it lies within a specific distance from it (Reachability), identifying core, border and noise points as the result of transitively connected points (Connectivity) [46]. More specifically, a core point is a point that has m points within a distance of n from itself, whilst a border point has at least one core point within the distance of n . All other points are considered as noise. The algorithm itself proceeds by randomly picking up a point from the dataset, that means, picking one node from the similarity graph, until every point was visited. All $minPoint$ -points within a radius of ϵ around the randomly chosen point are considered as one cluster. DBSCAN proceeds by recursively repeating the neighbourhood calculations for each neighbouring point, resulting in n clusters.

Spectral Clustering involves dimensionality reduction in advance before using standard clustering methods such as k -means. For dimensionality reduction, the similarity graph \mathcal{W} is transformed into the so-called graph Laplacian matrix L , which describes the relations of the nodes and edges of a graph, where the entries are defined by

$$L_{ij} := \begin{cases} \deg(z_i) & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and } w_{ij} > 0, \\ 0 & \text{else} \end{cases} \quad (16)$$

with $\deg(z_i) = \sum_{j=1}^{|\mathcal{D}(Z)|} w_{ij}$. For decorrelation, data in the graph Laplacian matrix L is decomposed into its sequence of eigenvalues and the corresponding eigenvectors. The eigenvectors form a new uncorrelated orthonormal basis and are thus suitable for standard clustering methods. The observations of the reduced data matrix whose columns contain the smallest k eigenvectors can now be clustered using k -means. An observation assigned to cluster C_i with $i = 1, \dots, k$ can then be traced back to its entity $z \in \mathcal{D}(Z)$ by indices.

We evaluate both clustering approaches as part of SA₄PG in Section 6.3. To improve comparability, we compare both clustering approaches as described in the next Section.

6.2.2 Clustering Comparison Approach

We compare DBSCAN and Spectral Clustering in MOP₄SA by identifying clusters with each clustering approach and use resulting clusters within SA₄PG respectively, i.e., run SA₄PG once using clusters determined by DBSCAN and once using clusters determined by Spectral Clustering.

Since DBSCAN is able to automatically determine the numbers of clusters, we use DBSCAN to identify same and provide the resulting number of clusters as a parameter when performing Spectral Clustering. As DBSCAN is capable to also classifies noise, i.e., entities, which cannot be assigned to a cluster, we use the number of points classified as noise plus the number of clusters as the number of total clusters in Spectral Clustering. Further, for DBSCAN we provide $minPoints = 2$ as the minimum number of entities in a cluster to allow for the maximum number of clusters in general. The eps parameter is automatically determined using the kneedle

algorithm [47]. For Spectral Clustering we just provide parameter k for the total number of clusters, which was previously determined by DBSCAN. Note that the total number of clusters n , which was determined by DBSCAN, does not necessarily corresponds to the best number of clusters for Spectral Clustering. Nevertheless, we obtain results that show which algorithm, given the same input n , is better at separating entities in the multidimensional space.

In the following, we perform a detailed comparison between the two clustering approaches as part of SA₄PG with different parameters for the symbolisation scheme in MOP₄SA using the approach to compare the two clustering mechanisms as described here.

6.3 Preventing Groundings

MOP₄SA is affected by (a) the efficiency of the clustering algorithm used, (b) the similarity measure itself, (c) and its hyper parameters such as order d , delay τ and δ for the arithmetic mean as defined in Eq. (9). We evaluate MOP₄SA as part of SA₄PG. Specifically, we approximate symmetry clusters using MOP₄SA with different settings and (i) perform inference using the the symmetry clusters to prevent the model from grounding, and (ii) compare it with exact lifted inference and calculate Kullback Leibler divergence (KLD) between query result to determine the error introduced through SA₄PG. A KLD with $D_{KL} = 0$ indicates that both distributions are equal. Inference in DPRMs is performed by the lifted dynamic junction tree algorithm. Details can be found in [5].

We ran 54 experiments in total with different parameter combinations $d \in \{2, 3, 4\}$, $\tau \in \{1, 2, 3\}$, $\delta \in \{0.05, 0.1, 0.15\}$ and clustering through Spectral Clustering and DBSCAN. For comparison, we perform query answering given sets of evidence, i.e., we perform inference by answering the prediction query $P(Supply_i(Z), Idle_i(Z))$ for each time step $t \in \{4, \dots, 51\}$ and obtain a marginal distribution for each entity $z \in \mathcal{D}(Z)$. We repeat query answering three times, once without preventing any groundings, once with preventing groundings using the clusters determined by DBSCAN, and once again with preventing groundings but using clusters determined by Spectral Clustering for each parameter combinations. Note that we only discuss results for a sub-selection of the parameter combinations, which give good results in terms of accuracy in inference under preventing groundings, while Table 1 and Table 2 at the end of this paper show the full results for all parameter combinations. Table 1 and Table 2 show results for time intervals $t \in \{\{5, 10\}, \{10, 15\}, \{15, 20\}, \{20, 25\}, \{25, 30\}, \{30, 35\}, \{35, 40\}, \{40, 45\}, \{45, 50\}\}$.

We evaluate runtime in seconds s , the number of groundings $\#_{gr}$ and KLD D_{KL} . Note that, $\#_{gr}$ shows the number of clusters after time t , while n shows the initial number of clusters. Thus, the number of additional groundings at a specific timestep equals to $\#_{gr}$ minus n . Preventing groundings aims at keeping a lifted model as long as possible. A basic prerequisite for this is that similarities exists in the data. As to that, the variable $n_{\geq 1}$ shows the number of initial clusters, which contain more than one entity directly after clustering, i.e., clusters in which similarly behaving entities have been arranged. Note that similar to n , $n_{\geq 1}$ does not change over time. With increasing order d the number of neighbouring data points are increasing, i.e., the classification contains more long term patterns. With increasing delay τ , long-term behaviour is extended even further, while also allowing for temporary deviations. For data

range dependence, in similarity counting we test different delta δ_{\leq} .

The number of clusters with more than one entity $n_{\geq 1}$ relative to the total number of clusters n are important in evaluating how well symmetries are exploited. When n is small, i.e. when n is significantly smaller than the total number of entities $|\mathcal{D}(Z)|$, a value of $n_{\geq 1}$ close to n is desirable since it indicates that many entities show symmetries with each other. If $n_{\geq 1}$ is significantly smaller than n , then only a few entities show symmetries, which on the one hand leads to a better accuracy in the inference since many entities are considered at a ground level, but on the other hand runtime will suffer greatly. As to that, Figure 6 shows a comparison for different parameter combinations and clustering approaches. The red line denotes the total number of clusters n independently of the number of entities included in a cluster, while the bars only show the number of clusters with more than one entity $n_{\geq 1}$. Note that we also include entities, which are treated on a ground level already by the time after learning clusters, in the total number of cluster, i.e., clusters can also only include one entity. Since lifting highly depends on the degree of similarities, only those clusters with more than one entity are of interest. Each pair of bar plots correspond to a different experiment with different parameter combinations.

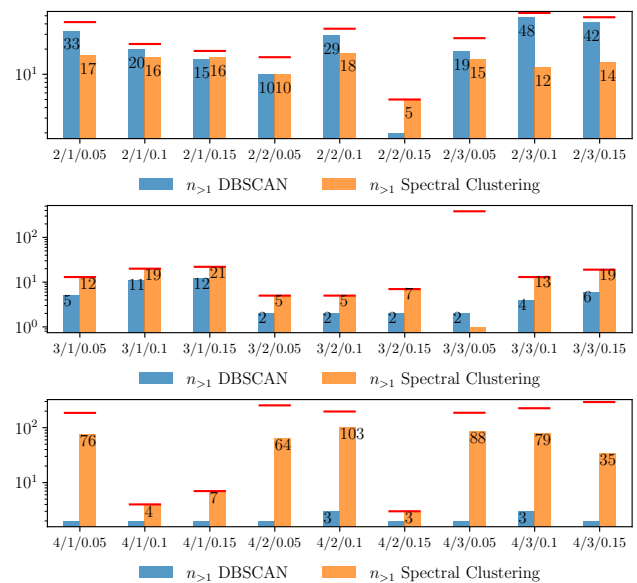
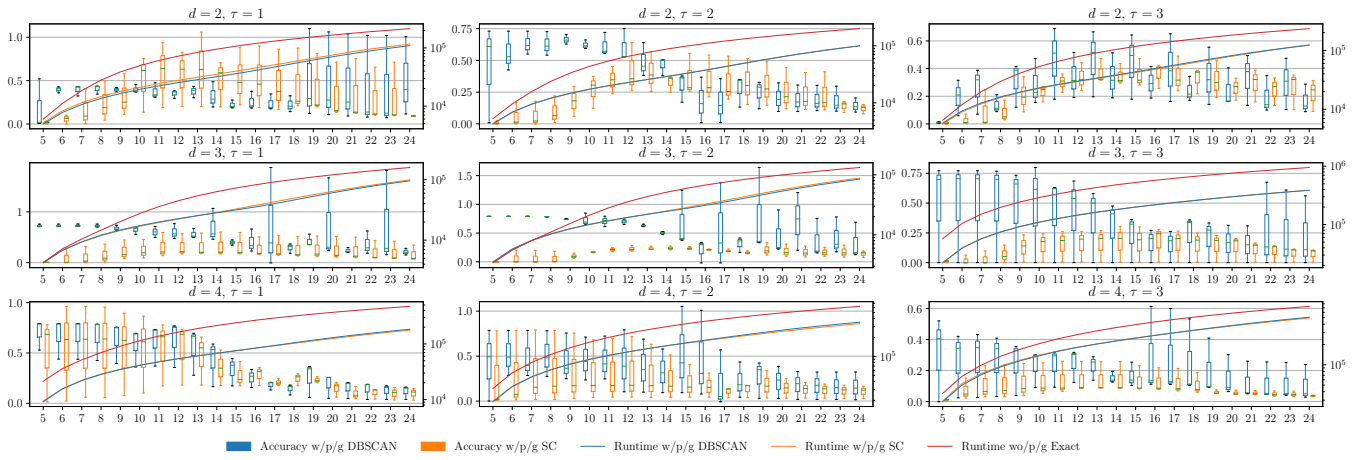
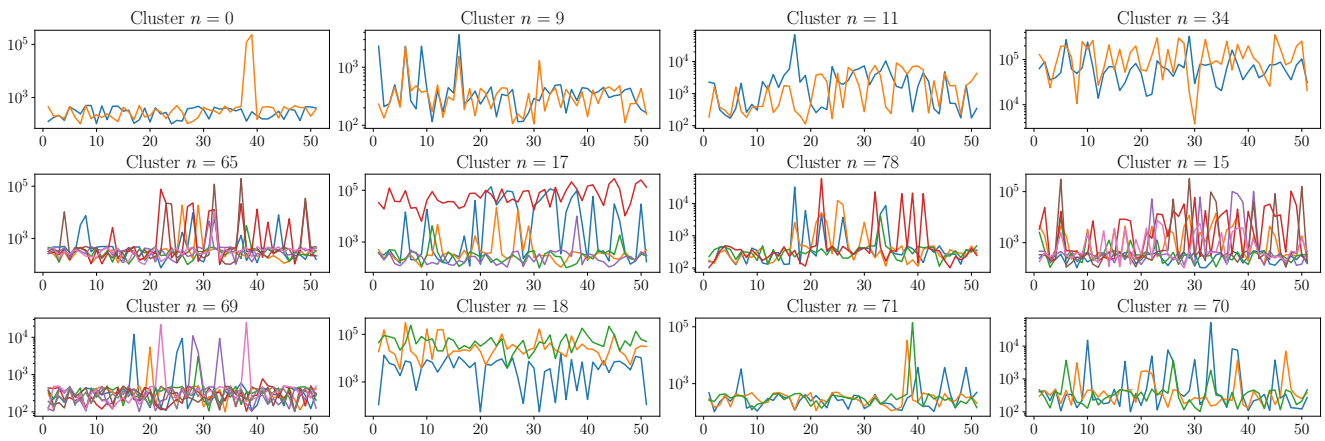


Figure 6: Comparison of number of clusters with more than one entity between DBSCAN and Spectral Clustering.

Due to space limitations we shorten order d , delay τ and delta δ_{leq} in the graph by the triple $d/\tau/\delta_{leq}$. From Fig. 6 it is most noticeable that the number of clusters with more than one entity $n_{\geq 1}$ for higher orders d is much less than for lower orders. This intuitively makes sense, since with higher orders d long term behaviour is captured much better than with lower orders and thus only a few clusters can be determined. For $d = 2$ much more clusters are identified since a smaller time span is considered resulting in a higher possibility of showing similarities. This observation also applies to increasing delays τ . The experiment with order $d = 3$, delay $\tau = 3$ and delta $\delta_{\leq} = 0.05$ is a good example for cases where not many similarities have been identified, but many entities are treated on a



(a) Accuracy and runtime in inference under SA₄PG for different parameter combinations.



(b) Supply over time t for a selection of clusters, which have been learned based on Supply/Idle time data for $0 < t \leq 4$ using Spectral Clustering.

Figure 7: Accuracy and runtime data based on query results under SA₄PG including raw supply data for entities within clusters.

ground level, i.e., accuracy will be good, but runtime will suffer. In the following, we look at accuracy results and will also come back to this example.

By comparing the KLD D_{KL} as a result of inference without preventing groundings and with preventing groundings based on clusters determined by DBSCAN and Spectral Clustering, it is noticeable that for clusters determined by Spectral Clustering in average a lower D_{KL} compared to DBSCAN results. This is explainable with better handling of higher dimensional data in Spectral Clustering. Figure 7a shows a comparison between the accuracy for both cases and different parameter combination. Each subplot corresponds to a different order d and delay τ , while the box-plot itself shows the variation of the accuracy over different deltas $\delta_{\leq} = \{0.05, 0.1, 0.15\}$ over time t . Note that we only plot data until $t = 24$ for better visibility and as the effect of any wrong evidence, which was brought in by preventing groundings, starts to level off. This happens since groundings are only prevented until the threshold H is reached, i.e., any other evidence afterwards at a later timestep balance out the effect of any wrong evidence at an early timestep after learning symmetry clusters. The blue box-plots in Fig. 7a correspond to the KLD D_{KL} with DBSCAN as the clustering approach in MOP₄SA, while the orange box-plots correspond to the KLD D_{KL} with on Spectral

Clustering as the clustering approach in MOP₄SA. The solid blue, orange and red line correspond to the runtime for answering a query for the specific time step. From the plots, we can see that for higher orders and delays, i.e., with increasing time spans each ordinal represents, that D_{KL} is decreasing. Considering the total number of clusters for each experiment (see Fig. 6), this follows as not many similarities can be found in the data, but more entities are handled on a ground level, i.e., increasing accuracy. On the other hand, runtime drastically increases as symmetries are no longer exploited. Compared to exact reasoning, runtime is noticeably smaller in inference under SA₄PG. To look again at the experiment with order $d = 3$, delay $\tau = 3$ (as highlighted above), the KLD D_{KL} is considerably small especially for Spectral clustering, but the runtime of the inference is very poor compared to all other experiments.

In SA₄PG, the violation threshold H is set to 5, i.e., groundings due to any inequalities are prevented for an entity H times. After $t = 10$ the number of groundings $\#_{gr}$ (see Table 1) are still the same as after learning the entity similarity cluster, i.e., all groundings are prevented in the initial timesteps after learning the clusters. Still, if entities behave similarity in early timesteps, the threshold H is reached far later in time. Thus, if in clustering based on the similarity graph

the entities with similarities are identified better, then groundings will occur much later in time. The longer D_{KL} stays small, the better cluster fit, i.e., the error introduced in inference through preventing groundings is kept small. In Fig. 7a we see that the accuracy suffers approximately for all experiments around $t = 10$, i.e., after 4 more timesteps after learning the clusters. Figure 7b depicts raw supply data for a selection of clusters as a result of running MOP₄SA based on data for $t = [0, 4)$ with parameters $d = 2$, $\tau = 1$ and $\delta = 0.05$ for symbolisation and Spectral Clustering. Even though only providing a small amount of training data, we can see that symmetrical behaviour continues for most of the clusters until approx. $t = 10$, like especially for clusters $n \in \{0, 11, 34, 65, 78, 69, 71\}$ and therefore support the insight, which we have got based on Fig. 7a. For simplicity only raw supply data is plotted even though symmetry clusters are determined based on supply/idle data.

The best results are achieved with Spectral Clustering as part of SA₄PG for the parameter combinations $d = 2$, $\tau = 1$, $\delta_{\leq} = 0.1$, $d = 3$, $\tau = 3$, $\delta_{\leq} = 0.1$ and $d = 3$, $\tau = 2$, $\delta_{\leq} = 0.05$, which we will also further refer to in the following Section. Generally, when reasoning under time constraints, preventing grounds is a reasonable approach as it prevents groundings in the long term and therefore speeds up inference.

Entity similarity can change over time, i.e., to further prevent the model from grounding it is beneficial to relearn symmetry structures at some time. In the following we propose MOP₄SCD and use it to identify points in time when relearning clusters is beneficial.

7 Multivariate Ordinal Pattern for Symmetry Change Detection (MOP₄SCD)

Symmetries in temporal models can change over time as already seen in Fig. 5. Therefore, symmetry cluster, after they have been learned, may only stay valid for a certain period of time. Further, some are valid for a longer period of time, some not. To identify points in time when relearning symmetry clusters is reasonable, we use the similarity graph as an intermediate output of running MOP₄SA and check if the similarity graph has changed *significantly*. More specifically, we continue running MOP₄SA for every timestep, but instead of for continuously relearning symmetry clusters, we prevent relearning clusters in MOP₄SA after the initial sync run until the graph has changed *significantly enough*. To identify such points in time with a significant change, we introduce MOP₄SCD taking as inputs a similarity graph for two consecutive timesteps and calculating a distance measure between both. In case the distance measure is above a certain threshold we consider those points as change points to trigger the cluster relearning process. MOP₄SCD is based on the assumption, that clusters no longer stay valid, if entities within a cluster no longer show the same similarity to its cluster entities as in the previous timesteps, i.e., the similarity counts is no longer proportionally scaling as before. Those entities might transition to another cluster, since its showing more similarity with another cluster. Informally, if the similarity graph changes over time in a *constant and balanced way*, symmetry clusters stay valid, but if the similarity graph changes over time in an *unbalanced manner*,

i.e., if similarity counts change significantly, there is a change in the structure of the symmetry clusters. To illustrate that, let us look at Figure 8. The Figure shows a similarity graph based on which two clusters have been identified.

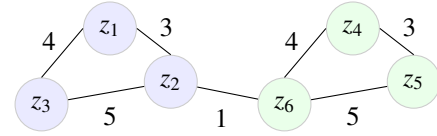


Figure 8: Overview of potential unbalanced changes in a similarity graph.

Nodes $S^1 = \{z_1, z_2, z_3\}$ (coloured in blue) denote a cluster S^1 and the nodes $S^2 = \{z_4, z_5, z_6\}$ (coloured in green) denote another cluster S^2 . Both clusters are connected through nodes z_2 and z_6 since for both a similarity was measured at any timestep before learning clusters. Relearning clusters becomes necessary if the cluster structure itself changes. This happens either

- if similarities between entities of different clusters changes, e.g., if the similarity between z_2 and z_6 increases and might require to merge the clusters or even split them into more than two clusters, which we denote as a *unbalanced interclusteral change*,
- or if similarities within a cluster change disproportionately, e.g., if similarities for S^2 changes only for a subset of the entities such as for z_4 and z_5 but not proportionally for all entities such as z_4 and z_6 and z_6 and z_5 requiring to split the cluster even further, which we denote as a *unbalanced intraclusteral change*.

As follows we define both unbalanced interclusteral and intraclusteral change measures and combine both into a distance measure denoting the unbalanced change between consecutive timesteps. Both unbalanced inter- and intraclusteral changes are determined based on the similarity graph \mathcal{W}^t from the current to the next time step \mathcal{W}^{t+1} under current symmetry clusters \mathcal{S} , with interclusteral changes defined as

$$d_{inter}(\mathcal{W}^t, \mathcal{W}^{t+1}, \mathcal{S}) = \sum_{\substack{S^i \in \mathcal{S} \\ S^j = en(S^i) \cap en(\mathcal{S})}} \frac{\sum_{\substack{i \in en(S^i) \\ j \in en(S^j)}} [w_{ij}^{t+1} = w_{ij}^t + 1]}{|en(S^i)| \cdot |en(S^j)|} \quad (17)$$

where $[x] = 1$ if x and, 0 otherwise for $en(S^i) \cap en(S^j) = \emptyset$ and intraclusteral changes defined as

$$d_{intra}(\mathcal{W}^t, \mathcal{W}^{t+1}, \mathcal{S}) = \sum_{S^i \in \mathcal{S}} \frac{\sum_{i, j \in en(S^i), i < j} [w_{ij}^{t+1} - w_{ij}^t = 0]}{|en(S^i)| \cdot |en(S^i)|} \quad (18)$$

where $[x] = 1$ if x and, 0. Both d_{inter} and d_{intra} are merged into one combined measure with

$$d(\mathcal{W}^t, \mathcal{W}^{t+1}, \mathcal{S}) = \frac{d_{inter} + d_{intra}}{|\mathcal{S}|} \quad (19)$$

Simply speaking, d_{inter} counts the number of increases in weights across different clusters S^i and S^j such as shown in Fig. 8 for entities z_2 and z_6 . The resulting count is normalised by dividing through the number of comparison between entity pairs of the clusters $en(S^i)$ and $en(S^j)$, resulting in measure between 0 and 1 with a value close

to 1 denoting a maximum dissimilarity. Similarly, d_{intra} counts the occurrences of no weight increases within entity pairs of a similar cluster S^i . To ensure that entities within a cluster continue to behave the same, weights should proportionally increase equally distributed within the cluster. If there is no increase in weights most likely the entities discontinue to behave similarly. The resulting count is equally normalised with a value close to 1 denoting a maximum dissimilarity. Finally, both d_{inter} and d_{intra} are combined in a single measure also count normalised to determine a distance measure between 0 and 1. If $d(\mathcal{W}^t, \mathcal{W}^{t+1}, \mathcal{S}) = 0$, the change in the similarity graph is balanced, if $d(\mathcal{W}^t, \mathcal{W}^{t+1}, \mathcal{S}) > 0$, it is unbalanced. If $d(\mathcal{W}^t, \mathcal{W}^{t+1}, \mathcal{S}) > b$, $b \in \mathbb{N}_{>0}$ it may be worthwhile to (re)perform clustering and (re)build symmetry clusters.

As follows, we evaluate MOP₄SCD based on clusters determined by MOP₄SA for the same parameters as in the experiments performed in Section 6.

8 MOP₄SCD in Application

We evaluate MOP₄SCD based on clusters determined using MOP₄SA as described in Section 6. Since Spectral Clustering works better than DBSCAN in identifying clusters, we here only use clusters determined by Spectral Clustering as part of MOP₄SA. We run experiments 27 experiments in total for the same parameter combinations $d \in \{2, 3, 4\}$, $\tau \in \{1, 2, 3\}$ and $\delta \in \{0.05, 0.1, 0.15\}$ as in Section 6. For each experiment we calculate $d(\mathcal{W}^t, \mathcal{W}^{t+1}, \mathcal{S})$ for timesteps $t = 5, \dots, 51$. Clusters are learned based on a similarity graph with data for $t = 1, \dots, 4$.

In this Section we discuss results for a sub-selection of the parameter combinations, which give good results in terms of accuracy in inference under preventing groundings as seen in Section 4, while Table 3 at the end of this paper shows detailed results for all parameter combinations. Each column in Table 3 shows the distance measure for consecutive timesteps, e.g., for $t = 5$, the distance is derived based on the similarity graph for timestep $t = 4$ to $t = 5$, i.e., $d(\mathcal{W}^4, \mathcal{W}^5, \mathcal{S})$. Note that since $d(\mathcal{W}^4, \mathcal{W}^5, \mathcal{S})$ is calculated for two consecutive timesteps, the distance measure has to be added up over time to derive the overall distance between more than two timesteps. Overall, the distance measure varies for different parameter combinations with in the optimal case showing an unbalanced change in weights of approximately 1.6% and in the worst case of approximately 22.4% between two consecutive timesteps. The distance measure is highly affected by the number of clusters n . In the case that the number of clusters with more than one entity $n_{>1}$ is considerably small compared to the total number of clusters n , the distance measure $d(\mathcal{W}^t, \mathcal{W}^{t+1}, \mathcal{S})$ is also considerably low since d_{inter} and d_{intra} , see Eq. (17) and Eq. (18), always return no unbalanced change for clusters with just one entity, i.e., for entities which are already being treated on a ground level. MOP₄SA aims at preventing groundings to speed up inference, i.e., lead to an increase in runtime. Therefore, choosing parameters d , τ and δ for MOP₄SA and consequently for MOP₄SCD is a trade-of between losses in accuracy and a speed up in inference.

The parameter combinations $d = 2, \tau = 1, \delta_{\leq} = 0.1$, $d = 3, \tau = 3, \delta_{\leq} = 0.1$ and $d = 3, \tau = 2, \delta_{\leq} = 0.05$ give good results in MOP₄SA as shown in Section 6. Results for MOP₄SCD also

support this. Figure 9 shows the KLD D_{KL} in conjunction with results from MOP₄SCD. Each subplot corresponds to a different parameter combination with the blue line corresponding to the KLD D_{KL} , the solid red line to the distance measure $d(\mathcal{W}^t, \mathcal{W}^{t+1}, \mathcal{S})$ for two consecutive timesteps t and $t + 1$ and the dashed red line for the cumulative distance measure, i.e., from $t = 0$ until the current timestep t . Note that the cumulative distance is log scaled and can be read of from the right y-axis. The highlighted red area in each subplots mark the interval when the cumulative distance measure becomes greater then 50% until it has reached 100%, i.e., with a change of 100% that all relations between all entities have been affected.

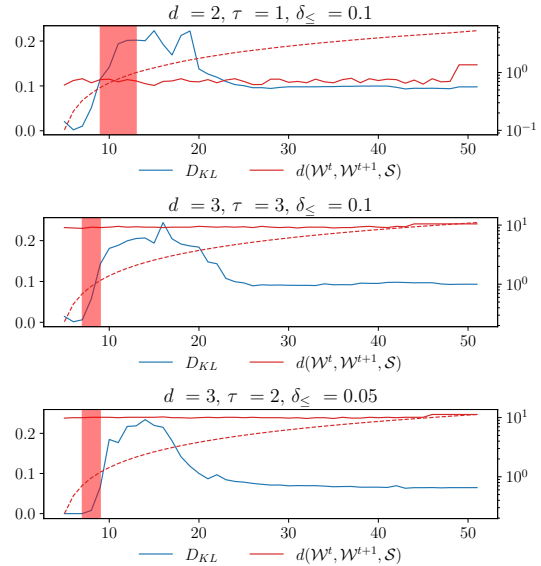


Figure 9: Results of MOP₄SCD for three parameter combinations which show the best results for MOP₄SA and SA₄PG. Further results can be found in the appendix.

Similar to the experiments in Section 6, clusters \mathcal{S} have been determined by MOP₄SA based on data for $0 > t \leq 4$. For all upcoming timesteps the clusters \mathcal{S} have been used to prevent groundings, i.e., execute SA₄PG as part of inference, see Algorithm 1. The KLD D_{KL} for all experiments as shown in Fig. 9 similarly raises up to a value of approx. 0.25 with its peak around $t = 15$. In contrast $d(\mathcal{W}^t, \mathcal{W}^{t+1}, \mathcal{S})$ varies across experiments and has for the experiment with parameter combination $d = 2, \tau = 1$ and $\delta_{\leq} 0.1$ its best value of approx. 0.1, i.e., an unbalanced change of approx. 10% over time. For the two other experiments $d(\mathcal{W}^t, \mathcal{W}^{t+1}, \mathcal{S})$ is with 0.22 similar. Correspondingly, the cumulative distance reaches a value of 0.5 at timestep $t = 9$ until it reaches a value of 1 at $t = 14$ for the first experiment, while for the two other experiments the cumulative distance reaches a value of 0.5 at $t = 7$ and a value of 1 already at $t = 9$. I.e., clusters \mathcal{S} are valid for a longer period of time using MOP₄SA with a parameter combination of $d = 2, \tau = 1, \delta_{\leq} = 0.1$. Further, for that parameter combination, D_{KL} settles off once a cumulative distance of 1 has reached. Settling off happens due to the amount of new evidence leading to more groundings removing the effect of any wrongly introduced evidence in previous timesteps. Relearning clusters at a threshold of 0.5 is here beneficial to prevent the D_{KL} from further increasing. That means relearning at $t = 9$, i.e., clusters are valid for approx. 4 timesteps after learning

them, which corresponds to a full month in our example application and therefore is a good result to considerably speed up inference while only introducing a small error in inference.

9 Conclusion and Future Work

Evidence lead to groundings in dynamic probabilistic relational models over time, negating runtime benefits in lifted inference. This paper provides MOP₄SA, SA₄PG and MOP₄SCD as a rich toolset to identify model symmetries as part of the model construction process, use those symmetries to maintain a lifted representation by preventing groundings a priori and detect changes in model symmetries after the model construction process. Preventing groundings a priori to maintain any lifted representation is important in lifted inference to preserve its runtime benefits. MOP₄SA detects symmetries across entities of the models domain using a multivariate ordinal pattern symbolisation approach and building a similarity graph for spectral clustering to identify sets of entities with symmetrical behaviour regarding a context of the model (symmetry clusters). Symmetry clusters are used in SA₄PG as part of query answering to prevent any unnecessary model splits by evidence, e.g., due to one time events. Symmetry structures can change over time, which MOP₄SCD detects based on the similarity graph, an intermediate output of MOP₄SA, and provide a distance measure denoting the degree of any unbalanced structural change to identify points in time when relearning symmetry clusters is beneficial.

The main contribution of this paper are the extension by theoretical and experimental results on the original papers [1, 2] and the introduction of MOP₄SCD as a mechanism to detect structural changes to complement MOP₄SA, SA₄PG as a rich toolset to prevent groundings a priori. We show, that MOP₄SA requires only a small amount of *training data* to come up with a good approximation of symmetry structures. Generally, MOP₄SA aims at determining symmetry structures which stay valid for shorter time periods. This follows, since MOP₄SA is not capable to capture any reoccurring patterns or periodicity, e.g., due to seasonality. MOP₄SA can be extended to capture such behaviour, but this would also increase the complexity of the overall approach. Due to this and since capturing symmetries for longer time spans, especially in real-world applications which normally change much faster, is not feasible, we focus with MOP₄SA as being a simple and easy to compute framework, requiring only few historical data points for learning, to identify symmetries for the short term future. In addition to MOP₄SA, MOP₄SCD supports in inference by identifying points in time when relearning clustering for SA₄PG is reasonable.

With preventing groundings a priori we complement existing approaches, which focus on retaining lifted representation after a model has already been splitted. In general, our approach works well with any other approach undoing splits after they occurred when moving forward in time, e.g., in message passing by merging sets of entities when those align again, denoted as *temporal approximate merging*, as proposed in [25]. Combining both kind of approaches brings together the best of both worlds: (a) While with *determining approximate model symmetries* a priori, we can use the full amount of historical training data to prevent groundings, (b) and with *temporal approximate merging*, we can merge non-preventable

parfactor splits even after they occurred, i.e., a posterior.

Since MOP₄SA is designed to work with small amounts of data to provide symmetry clusters very quickly for the short term future, the overhead MOP₄SA and MOP₄SCD bring into query answering need to be kept to a minimum. Applying the symbolisation scheme to identify symmetries is already a suitable mechanism, but with the clustering approach we still depend on existing approaches, which are considerably costly. The investigation of more performant clustering approaches, e.g., taking advantage of some sort of incremental changes to clustering after the initial learning step, are left for future work.

List of Symbols

R	set of random variables
L	set of logical variables
Φ	set of factor names
D	set of entities
$\mathcal{D}(L)$	domain of a logvar
$C, (X, C_X)$	constraint restricting logical variables
$A(L_1, \dots, L_n)$	parameterised logical variable (PRV)
$g, \phi(\mathcal{A})_C$	parfactor
$gr(P)$	grounding
$lv(P)$	logical variables
$\mathcal{R}(A)$	range of a PRV
G	model
G_t	local model
E	evidence, set of events
Q	query term
\mathcal{X}	multivariate time series
τ	delay between successive time points
d	order of ordinal pattern
w_{ij}	similarity count
\mathcal{W}	similarity graph
S	symmetry cluster
$en(S)$	objects in a symmetry cluster
S	set of symmetry clusters
P	parfactor partition
L	Laplacian matrix
D_{KL}	Kullback-Leibler divergence
δ_{\leq}	mean delta
$d(\mathcal{W}^t, \mathcal{W}^{t+1}, S)$	similarity change measure

References

- [1] N. Finke, M. Mohr, "A Priori Approximation of Symmetries in Dynamic Probabilistic Relational Models," in S. Edelkamp, R. Möller, E. Rueckert, editors, KI 2021: Advances in Artificial Intelligence, 309–323, Springer International Publishing, Cham, 2021.
- [2] N. Finke, R. Möller, M. Mohr, "Multivariate Ordinal Patterns for Symmetry Approximation in Dynamic Probabilistic Relational Models," in AI 2021: Advances in Artificial Intelligence - 34rd Australasian Joint Conference, Lecture Notes in Computer Science (LNCS), Springer International Publishing, In Press.
- [3] N. Finke, M. Gehrke, T. Braun, T. Potten, R. Möller, "Investigating Maturity of Probabilistic Graphical Models for Dry-Bulk Shipping," in M. Jaeger, T. D.

- Nielsen, editors, Proceedings of the 10th International Conference on Probabilistic Graphical Models, volume 138 of *Proceedings of Machine Learning Research*, 197–208, PMLR, 2020.
- [4] Y. Xiang, K.-L. Poh, “Time-Critical Dynamic Decision Making,” 2013.
- [5] M. Gehrke, T. Braun, R. Möller, “Lifted Dynamic Junction Tree Algorithm,” in Proceedings of the International Conference on Conceptual Structures, 55–69, Springer, 2018.
- [6] D. Poole, “First-order Probabilistic Inference,” in Proc. of the 18th International Joint Conference on Artificial Intelligence, 985–991, 2003.
- [7] D. Akyar, “The Effects of Global Economic Growth on Dry Bulk Shipping Markets and Freight Rates,” 2018.
- [8] Z. Wang, X. Wu, K. L. Lo, J. J. Mi, “Assessing the management efficiency of shipping company from a congestion perspective: A case study of Hapag-Lloyd,” *Ocean & Coastal Management*, **209**, 105617, 2021, doi: <https://doi.org/10.1016/j.ocecoaman.2021.105617>.
- [9] C. Jiang, Y. Wan, A. Zhang, “Internalization of port congestion: strategic effect behind shipping line delays and implications for terminal charges and investment,” *Maritime Policy & Management*, **44**(1), 112–130, 2017, doi: 10.1080/03088839.2016.1237783.
- [10] T. Notteboom, “The Time Factor in Liner Shipping Services,” *Maritime Economics and Logistics*, **8**, 19–39, 2006, doi:10.1057/palgrave.mel.9100148.
- [11] M. Niepert, G. Van den Broeck, “Tractability through Exchangeability: A New Perspective on Efficient Probabilistic Inference,” in AAAI-14 Proceedings of the 28th AAAI Conference on Artificial Intelligence, 2467–2475, AAAI Press, 2014.
- [12] G. V. den Broeck, M. Niepert, “Lifted Probabilistic Inference for Asymmetric Graphical Models,” *CoRR*, **abs/1412.0315**, 2014.
- [13] N. Taghipour, D. Fierens, J. Davis, H. Blockeel, “Lifted Variable Elimination: Decoupling the Operators from the Constraint Language,” *Journal of Artificial Intelligence Research*, **47**(1), 393–439, 2013.
- [14] K. Kersting, “Lifted Probabilistic Inference,” in Proceedings of the 20th European Conference on Artificial Intelligence, ECAI’12, 33–38, IOS Press, NLD, 2012.
- [15] G. Van den Broeck, A. Darwiche, “On the Complexity and Approximation of Binary Evidence in Lifted Inference,” in C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26, Curran Associates, Inc., 2013.
- [16] P. Singla, A. Nath, P. Domingos, “Approximate Lifting Techniques for Belief Propagation,” in Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, AAAI’14, 2497–2504, AAAI Press, 2014.
- [17] P. Singla, P. Domingos, “Lifted First-Order Belief Propagation,” in Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 2, AAAI’08, 1094–1099, AAAI Press, 2008.
- [18] B. Ahmadi, K. Kersting, M. Mladenov, S. Natarajan, “Exploiting symmetries for scaling loopy belief propagation and relational training,” *Machine Learning*, **92**, 91–132, 2013.
- [19] C. Sutton, A. McCallum, “Piecewise Training for Structured Prediction,” *Machine Learning*, **77**, 165–194, 2009, doi:10.1007/s10994-009-5112-z.
- [20] D. Venugopal, V. Gogate, “Evidence-Based Clustering for Scalable Inference in Markov Logic,” in T. Calders, F. Esposito, E. Hüllermeier, R. Meo, editors, *Machine Learning and Knowledge Discovery in Databases*, 258–273, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.
- [21] P. Miettinen, T. Mielikäinen, A. Gionis, G. Das, H. Mannila, “The Discrete Basis Problem,” in J. Fürnkranz, T. Scheffer, M. Spiliopoulou, editors, *Knowledge Discovery in Databases: PKDD 2006*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- [22] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, “Equation of State Calculations by Fast Computing Machines,” *The Journal of Chemical Physics*, **21**(6), 1087–1092, 1953.
- [23] W. K. Hastings, “Monte Carlo Sampling Methods Using Markov Chains and Their Applications,” *Biometrika*, **57**(1), 97–109, 1970.
- [24] A. Nath, P. Domingos, “Efficient Lifting for Online Probabilistic Inference,” volume 2, 2010.
- [25] M. Gehrke, R. Möller, T. Braun, “Taming Reasoning in Temporal Probabilistic Relational Models,” in Proceedings of the 24th European Conference on Artificial Intelligence (ECAI 2020), 2020, doi:10.3233/FAIA200395.
- [26] R. Agrawal, C. Faloutsos, A. Swami, “Efficient similarity search in sequence databases,” in *Lecture Notes in Computer Science*, volume 730, Springer Verlag, 1993.
- [27] E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrotra, “Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases,” in *Knowledge and Information Systems*, 263–286, 2001, doi:10.1021/acsami.7b03579.
- [28] S. Kramer, “A Brief History of Learning Symbolic Higher-Level Representations from Data (And a Curious Look Forward),” in Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20, 4868–4876, 2020.
- [29] J. B. Kruskal, M. Liberman, “The Symmetric Time Warping Problem: From Continuous to Discrete,” in *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, Addison-Wesley Publishing Co., 1983.
- [30] C.-C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. Silva, A. Mueen, E. Keogh, “Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets,” in 2016 IEEE 16th International Conference on Data Mining (ICDM), 1317–1322, 2016.
- [31] F. Petitjean, J. Inglada, P. Gancarski, “Satellite Image Time Series Analysis Under Time Warping,” *IEEE Transactions on Geoscience and Remote Sensing*, **50**(8), 2012.
- [32] S. Salvador, P. Chan, “FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space,” 70–80, 2004.
- [33] D. F. Silva, G. E. A. P. A. Batista, “Speeding Up All-Pairwise Dynamic Time Warping Matrix Calculation,” in Proceedings of the 2016 SIAM International Conference on Data Mining, 837–845, Society for Industrial and Applied Mathematics, 2016.
- [34] B. Chiu, E. Keogh, S. Lonardi, “Probabilistic Discovery of Time Series Motifs,” in Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 493–498, 2003.
- [35] C. Bandt, B. Pompe, “Permutation Entropy: A Natural Complexity Measure for Time Series,” *Physical Review Letters*, **88**(17), 4, 2002.
- [36] M. Mohr, F. Wilhelm, M. Hartwig, R. Möller, K. Keller, “New Approaches in Ordinal Pattern Representations for Multivariate Time Series,” in Proceedings of the 33rd International Florida Artificial Intelligence Research Society Conference, 2020.
- [37] K. Keller, T. Mangold, I. Stolz, J. Werner, “Permutation Entropy: New Ideas and Challenges,” *Entropy*, **19**(3), 2017.
- [38] A. B. Piek, I. Stolz, K. Keller, “Algorithmics, Possibilities and Limits of Ordinal Pattern Based Entropies,” *Entropy*, **21**(6), 2019.
- [39] K. Keller, S. Maksymenko, I. Stolz, “Entropy Determination Based on the Ordinal Structure of a Dynamical System,” *Discrete and Continuous Dynamical Systems - Series B*, **20**(10), 3507–3524, 2015.
- [40] I. Stolz, K. Keller, “A General Symbolic Approach to Kolmogorov-Sinai Entropy,” *Entropy*, **19**(12), 2017.
- [41] A. Antoniouk, K. Keller, S. Maksymenko, “Kolmogorov-Sinai entropy via separation properties of order-generated σ -algebras,” *Discrete & Continuous Dynamical Systems*, **34**(5), 1793–1809, 2014.

- [42] K. Keller, "Permutations and the Kolmogorov-Sinai Entropy," *Discrete & Continuous Dynamical Systems*, **32**(3), 891–900, 2012.
- [43] D. Yang, L. Wu, S. Wang, H. Jia, K. X. Li, "How big data enriches maritime research – a critical review of Automatic Identification System (AIS) data applications," *Transport Reviews*, **39**(6), 755–773, 2019, doi:10.1080/01441647.2019.1649315.
- [44] R. Bellman, *Adaptive control processes: A guided tour*, Princeton legacy library, Princeton University Press, 2015.
- [45] A. L. Bertozzi, E. Merkurjev, "Chapter 12 - Graph-based optimization approaches for machine learning, uncertainty quantification and networks," in R. Kimmel, X.-C. Tai, editors, *Processing, Analyzing and Learning of Images, Shapes, and Forms: Part 2*, volume 20 of *Handbook of Numerical Analysis*, 503–531, Elsevier, 2019.
- [46] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, X. Xu, "DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN," *ACM Trans. Database Syst.*, **42**(3), 2017, doi:10.1145/3068335.
- [47] V. Satopaa, J. Albrecht, D. Irwin, B. Raghavan, "Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior," in 2011 31st International Conference on Distributed Computing Systems Workshops, 166–171, 2011, doi:10.1109/ICDCSW.2011.20.

Table 1: Accuracy scores of MOP4SA in SA4PG. Further results can be found in the appendix

d	τ	δ _≤	n	n>1	DBSCAN			Spectral Clustering				Exact		DBSCAN			Spectral Clustering			Exact	
					# _{gr}	s	D _{KL}	n>1	# _{gr}	s	D _{KL}	# _{gr}	s	# _{gr}	s	D _{KL}	# _{gr}	s	D _{KL}	# _{gr}	s
t = [05, 10]																					
1	1	0.05	42	33	42	19.6	0.342	17	42	19.1	0.317	306	38.7	66	37.8	0.294	81	37.5	0.621	357	95.2
		0.1	23	20	23	15.3	0.281	16	23	19.8	0.036	310	29.8	54	29.9	0.397	64	39.6	0.185	357	78.1
		0.15	19	15	19	14.5	0.441	16	19	14.5	0.126	268	25.1	46	28.4	0.401	64	28.8	0.857	351	69.6
2	2	0.05	16	10	16	14.1	0.578	10	16	13.9	0.172	251	22.8	47	27.6	0.535	56	27.7	0.388	348	65.7
		0.1	35	29	35	17.6	0.458	18	35	17.5	0.062	310	33.2	75	34.9	0.487	69	34.4	0.444	359	87.5
		0.15	5	2	5	12.5	0.724	5	5	12.4	0.014	180	15.8	29	24.4	0.655	48	24.8	0.224	342	50.9
3	3	0.05	27	19	27	15.7	0.240	15	27	15.7	0.026	300	29.1	71	31.3	0.442	57	30.8	0.249	358	78.7
		0.1	54	48	54	22.0	0.236	12	54	21.8	0.093	317	41.1	81	42.5	0.506	85	42.4	0.307	361	103.9
		0.15	48	42	48	21.2	0.070	14	48	20.2	0.129	309	37.7	80	41.0	0.182	70	39.0	0.392	355	96.5
1	1	0.05	13	5	13	13.6	0.756	12	13	13.6	0.021	177	17.1	31	26.5	0.682	63	27.3	0.152	339	55.8
		0.1	20	11	20	15.5	0.697	19	20	14.5	0.283	194	20.0	39	29.5	0.567	64	28.8	0.561	342	62.0
		0.15	22	12	22	15.2	0.706	21	22	15.1	0.058	195	20.3	40	29.3	0.620	67	30.2	0.207	344	63.7
3	2	0.05	5	2	5	12.6	0.782	5	5	12.4	0.011	165	15.1	28	24.4	0.667	45	24.7	0.205	338	50.1
		0.1	5	2	5	12.5	0.781	5	5	12.4	0.147	165	15.1	28	24.4	0.697	51	24.9	0.197	338	50.0
		0.15	7	2	7	12.8	0.776	7	7	12.7	0.019	168	15.4	25	24.9	0.599	48	25.2	0.246	340	51.3
3	3	0.05	383	2	383	249.7	0.000	1	383	247.8	0.001	385	244.0	384	488.7	0.000	383	485.8	0.005	385	480.0
		0.1	13	4	13	13.6	0.763	13	13	13.7	0.040	174	16.8	36	26.4	0.574	60	27.3	0.194	338	55.3
		0.15	19	6	19	14.8	0.694	19	19	14.5	0.111	199	19.8	39	28.6	0.597	64	28.8	0.267	340	61.8
1	1	0.05	186	2	186	79.0	0.445	76	186	79.9	0.047	296	90.9	205	154.3	0.306	203	155.1	0.163	358	214.2
		0.1	4	2	4	12.5	0.784	4	4	12.6	0.911	165	15.0	27	24.3	0.676	23	24.5	0.712	338	49.4
		0.15	7	2	7	12.8	0.779	7	7	12.8	0.644	167	15.4	25	24.9	0.694	38	25.0	0.571	338	51.1
4	2	0.05	255	2	255	127.3	0.219	64	255	127.9	0.018	336	137.4	277	251.4	0.155	261	249.1	0.043	366	297.7
		0.1	197	3	197	86.1	0.445	103	197	86.9	0.116	300	97.4	213	168.1	0.381	209	169.1	0.171	363	228.5
		0.15	3	2	3	12.5	0.779	3	3	12.4	0.779	165	14.9	21	24.2	0.699	27	24.2	0.647	338	48.9
3	3	0.05	187	2	187	79.7	0.427	88	187	80.6	0.027	293	89.1	199	155.4	0.277	196	156.3	0.076	360	212.1
		0.1	226	3	226	105.5	0.353	79	226	107.6	0.056	315	114.9	240	205.7	0.263	235	208.1	0.088	371	261.3
		0.15	293	2	293	158.5	0.024	35	293	164.7	0.173	362	169.2	299	315.9	0.116	301	316.1	0.256	384	356.4
t = [15, 20]																					
1	1	0.05	42	33	231	71.7	0.170	17	232	73.7	0.355	359	154.5	302	123.9	0.150	316	133.1	0.117	362	214.0
		0.1	23	20	229	58.9	0.324	16	230	69.4	0.201	359	127.9	312	104.8	0.182	318	115.3	0.116	362	178.4
		0.15	19	15	236	55.2	0.374	16	227	56.3	0.867	353	116.6	305	98.9	1.029	314	100.4	0.584	356	164.2
2	2	0.05	16	10	242	55.1	0.145	10	238	55.3	0.280	351	111.2	301	97.6	0.098	311	98.5	0.173	353	157.0
		0.1	35	29	230	67.9	0.354	18	214	64.9	0.527	359	143.1	305	117.1	0.286	317	114.6	0.313	362	199.6
		0.15	5	2	227	46.7	0.203	5	229	49.4	0.193	348	90.1	301	83.8	0.162	308	87.3	0.102	350	129.9
3	3	0.05	27	19	251	63.8	0.471	15	224	58.6	0.344	360	130.0	305	111.3	0.280	313	104.7	0.324	362	182.2
		0.1	54	48	211	78.1	0.392	12	239	80.5	0.237	362	168.6	298	131.4	0.346	330	139.4	0.132	364	233.9
		0.15	48	42	217	75.5	0.168	14	214	71.3	0.443	357	160.2	293	127.5	0.118	315	125.5	0.305	359	221.3
1	1	0.05	13	5	222	49.6	0.264	12	219	53.8	0.147	345	98.8	303	90.3	0.187	307	94.7	0.091	348	142.5
		0.1	20	11	222	54.6	0.348	19	224	57.1	0.535	346	108.3	297	98.0	0.291	309	100.4	0.446	349	155.5
		0.15	22	12	225	55.0	0.753	21	222	58.8	0.176	347	111.1	300	99.4	0.908	303	103.0	0.110	351	159.5
3	2	0.05	5	2	233	46.8	0.166	5	248	49.2	0.174	344	89.2	301	84.6	0.193	309	88.1	0.087	347	128.9
		0.1	5	2	230	46.8	0.358	5	231	49.4	0.216	344	89.0	300	84.5	0.774	297	86.9	0.255	347	128.8
		0.15	7	2	217	46.3	0.995	7	228	49.8	0.217	345	91.5	305	84.8	0.402	319	89.4	0.161	348	132.3
3	3	0.05	383	2	384	728.7	0.000	1	387	726.9	0.004	385	714.6	385	968.8	0.000	388	968.8	0.004	385	950.4
		0.1	13	4	234	51.0	0.309	13	236	55.0	0.202	344	98.2	301	91.9	0.493	310	96.3	0.135	347	141.8
		0.15	19	6	235	54.2	0.272	19	217	56.6	0.264	345	107.6	313	98.9	0.135	301	99.4	0.144	348	154.4
1	1	0.05	186	2	319	257.8	0.121	76	269	245.3	0.181	361	343.5	337	382.2	0.043	329	361.1	0.128	364	472.5
		0.1	4	2	231	46.6	0.272	4	211	45.3	0.307	344	88.0	300	83.8	0.195	301	82.0	0.043	347	127.2
		0.15	7	2	217	46.3	0.307	7	229	49.6	0.257	344	93.4	304	84.7	0.192	295	87.4	0.156	347	134.2
4	2	0.05	255	2	344	401.4	0.059	64	309	382.2	0.039	367	460.5	352	562.2	0.019	341	533.9	0.032	368	624.7
		0.1	197	3	324	275.1	0.697	103	260	264.5	0.158	365	362.6	346	406.3	0.336	324	381.1	0.134	366	497.2
		0.15	3	2	215	45.0	0.251	3	225	45.8	0.273	344	87.1	303	81.6	0.179	297	82.3	0.197	347	126.1
3	3	0.05	187	2	315	253.7	0.482	88	267	248.7	0.068	362	340.1	352	382.7	0.261	319	361.2	0.038	363	468.2
		0.1	226	3	324	325.9	0.104	79	290	321.7	0.082	373	413.8	357	472.8	0.052	339	460.3	0.063	373	564.5
		0.15	293	2	341	480.2	0.122	35	333	478.9	0.180	384	545.1	377	663.8	0.053	357	656.8	0.045	385	733.7

Table 2: Results of MOP₄SA for further timesteps with $t \geq 25$

d	τ	δ_{\leq}	n	DBSCAN			Spectral Clustering			Exact		DBSCAN			Spectral Clustering			Exact		
				$n_{>1}$	$\#_{gr}$	s	D_{KL}	$n_{>1}$	$\#_{gr}$	s	D_{KL}	$\#_{gr}$	s	$\#_{gr}$	s	D_{KL}	$\#_{gr}$	s	D_{KL}	$\#_{gr}$
t = [25, 30)																				
1	0.05	42	33	328	182.7	0.371	17	338	193.2	0.078	363	274.0	334	248.7	0.356	345	255.4	0.073	366	334.9
	0.1	23	20	328	155.2	0.102	16	340	167.2	0.093	362	229.2	338	207.6	0.088	341	220.7	0.094	364	280.7
	0.15	19	15	318	146.5	0.988	16	332	150.4	0.064	356	212.3	322	195.5	0.442	338	201.1	0.052	358	261.9
2	0.05	16	10	318	143.7	0.068	10	333	146.2	0.128	355	203.5	322	191.5	0.052	340	195.7	0.110	358	250.6
	0.1	35	29	324	172.1	0.105	18	337	171.8	0.076	362	256.8	334	229.4	0.076	341	230.3	0.064	364	314.5
	0.15	5	2	312	124.5	0.127	5	320	128.8	0.070	352	170.1	314	167.1	0.136	325	171.7	0.068	355	210.8
3	0.05	27	19	322	162.2	0.324	15	338	157.2	0.292	362	234.5	328	215.1	0.086	348	212.4	0.178	365	287.2
	0.1	54	48	331	194.4	0.233	12	348	205.7	0.070	364	299.1	342	261.1	0.119	352	275.6	0.065	366	365.4
	0.15	48	42	327	187.5	0.083	14	341	188.6	0.124	359	282.9	336	250.5	0.072	349	253.1	0.064	362	349.6
1	0.05	13	5	313	134.9	0.143	12	323	140.1	0.064	349	187.2	317	180.7	0.134	329	187.2	0.063	352	232.2
	0.1	20	11	313	145.7	0.494	19	332	149.9	0.169	351	203.2	320	194.9	0.440	338	200.9	0.097	354	251.7
	0.15	22	12	314	148.0	0.656	21	326	152.9	0.079	352	208.4	319	198.6	0.189	330	204.3	0.077	355	258.0
3	0.05	5	2	311	125.5	0.149	5	320	131.4	0.071	349	169.9	313	167.4	0.145	325	174.4	0.066	352	211.0
	0.1	5	2	310	125.4	0.631	5	313	127.8	0.145	349	169.0	312	167.1	0.590	322	170.2	0.129	352	210.0
	0.15	7	2	311	126.6	0.142	7	335	133.6	0.112	350	173.6	316	169.5	0.122	340	179.1	0.092	353	215.5
3	0.05	383	2	385	1214.0	0.000	1	388	1210.5	0.004	385	1186.4	385	1458.9	0.000	388	1454.6	0.004	385	1422.8
	0.1	13	4	309	136.1	0.523	13	322	142.5	0.090	349	186.0	312	181.3	0.492	327	189.5	0.089	352	231.0
	0.15	19	6	320	147.0	0.108	19	328	164.4	0.089	350	201.5	322	196.2	0.109	337	226.8	0.077	353	249.3
1	0.05	186	2	341	508.9	0.037	76	360	490.5	0.108	364	602.1	342	635.8	0.036	372	627.3	0.042	367	732.7
	0.1	4	2	310	124.1	0.148	4	311	122.5	0.503	349	167.1	312	165.6	0.139	317	164.3	0.524	352	207.6
	0.15	7	2	312	126.6	0.143	7	313	129.2	0.105	349	176.1	317	169.8	0.135	319	172.4	0.075	352	218.2
4	0.05	255	2	355	725.2	0.015	64	361	698.5	0.025	369	789.8	356	889.8	0.016	368	873.8	0.021	370	956.1
	0.1	197	3	352	540.4	0.313	103	363	516.2	0.115	367	632.7	353	676.1	0.303	368	658.3	0.102	369	772.0
	0.15	3	2	312	122.2	0.134	3	309	122.6	0.156	349	165.4	317	163.8	0.131	311	163.8	0.152	352	205.5
3	0.05	187	2	354	513.2	0.252	88	346	487.1	0.025	363	597.6	355	652.6	0.254	358	619.3	0.021	365	734.7
	0.1	226	3	360	623.2	0.050	79	361	611.6	0.030	373	715.9	362	775.4	0.047	368	769.2	0.028	375	868.0
	0.15	293	2	384	857.6	0.019	35	369	842.5	0.038	385	924.6	385	1054.2	0.013	374	1031.6	0.032	385	1115.7
t = [35, 40)																				
1	0.05	42	33	340	314.0	0.341	17	350	318.9	0.072	369	396.8	346	378.5	0.330	353	383.4	0.070	371	459.1
	0.1	23	20	346	261.6	0.078	16	342	274.7	0.095	367	333.0	349	316.7	0.074	349	329.3	0.093	370	385.8
	0.15	19	15	329	247.2	0.066	16	339	252.7	0.050	361	311.4	335	298.3	0.065	345	305.2	0.048	363	361.3
2	0.05	16	10	327	240.0	0.054	10	343	246.1	0.104	361	298.6	333	289.6	0.052	350	303.8	0.109	363	347.6
	0.1	35	29	338	288.0	0.070	18	345	290.0	0.063	367	373.8	343	347.7	0.069	348	350.8	0.064	369	432.8
	0.15	5	2	320	209.7	0.136	5	330	215.4	0.068	358	252.4	330	253.1	0.124	338	260.1	0.068	360	294.5
3	0.05	27	19	334	269.6	0.081	15	355	269.0	0.087	367	341.2	340	324.8	0.078	359	326.6	0.082	370	395.3
	0.1	54	48	347	329.5	0.107	12	353	344.8	0.064	369	438.8	352	398.8	0.100	360	415.8	0.063	373	506.7
	0.15	48	42	346	316.1	0.055	14	349	319.3	0.061	365	414.6	353	383.1	0.052	352	385.6	0.063	368	479.4
1	0.05	13	5	325	227.6	0.131	12	333	241.8	0.062	355	277.9	331	275.4	0.121	341	290.8	0.062	357	324.1
	0.1	20	11	325	245.1	0.416	19	343	253.5	0.066	357	300.9	333	296.5	0.386	347	306.7	0.065	359	350.9
	0.15	22	12	325	249.6	0.093	21	336	256.7	0.074	358	308.3	335	302.0	0.060	346	310.2	0.072	360	359.3
3	0.05	5	2	318	210.5	0.140	5	331	218.4	0.064	355	253.1	327	254.1	0.131	340	263.3	0.062	357	295.6
	0.1	5	2	318	209.7	0.568	5	326	213.6	0.132	355	251.9	328	253.4	0.527	332	257.9	0.134	357	298.2
	0.15	7	2	322	213.3	0.113	7	344	227.6	0.094	356	259.6	331	258.3	0.108	347	286.1	0.095	358	302.9
3	0.05	383	2	385	1701.7	0.000	1	389	1699.4	0.003	385	1659.9	385	1944.8	0.000	390	1943.8	0.003	385	1898.4
	0.1	13	4	317	227.3	0.472	13	332	237.5	0.091	355	276.8	327	274.4	0.443	342	286.3	0.095	357	323.4
	0.15	19	6	328	246.2	0.111	19	341	285.6	0.077	356	299.6	335	298.8	0.101	349	338.4	0.075	358	349.2
1	0.05	186	2	345	764.2	0.031	76	374	766.0	0.023	367	864.1	347	893.2	0.031	376	913.4	0.022	369	1003.2
	0.1	4	2	317	207.7	0.134	4	323	207.2	0.519	355	248.8	326	250.8	0.125	332	251.0	0.465	357	290.7
	0.15	7	2	323	213.8	0.134	7	323	216.7	0.076	355	261.9	331	259.1	0.123	330	261.7	0.149	357	305.5
4	0.05	255	2	356	1055.1	0.016	64	372	1045.3	0.021	370	1124.5	359	1221.2	0.015	376	1218.6	0.021	371	1291.4
	0.1	197	3	355	812.6	0.295	103	371	804.4	0.098	370	914.6	358	951.2	0.273	373	948.6	0.097	372	1067.9
	0.15	3	2	323	206.2	0.129	3	317	205.7	0.147	355	246.3	332	249.8	0.117	327	248.9	0.137	357	287.7
3	0.05	187	2	356	785.1	0.245	88	360	753.9	0.020	368	865.9	358	917.9	0.238	366	889.9	0.019	368	998.0
	0.1	226	3	364	929.7	0.046	79	373	929.2	0.028	377	1021.8	366	1084.0	0.045	376	1090.0	0.027	378	1175.8
	0.15	293	2	385	1251.4	0.012	35	377	1223.1	0.030	385	1304.6	385	1447.7	0.012	379	1416.0	0.028	385	1494.3

Table 3: Distances as a result of running MOP₄SCD between consecutive timesteps

d	τ	δ_{\leq}	$d(W^r, W^{r+1}, S)$																			
			5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
1	1	0.05	.067	.067	.068	.065	.066	.068	.067	.066	.063	.065	.069	.068	.062	.061	.064	.069	.064	.067	.068	
		0.1	.102	.112	.116	.107	.114	.115	.109	.114	.111	.105	.101	.11	.111	.116	.11	.109	.114	.107	.113	
		0.15	.135	.132	.142	.125	.136	.142	.142	.138	.14	.135	.137	.133	.139	.132	.129	.135	.131	.134	.143	
2	2	0.05	.105	.113	.114	.116	.113	.115	.113	.114	.115	.115	.116	.113	.117	.114	.115	.112	.113	.113	.116	
		0.1	.068	.068	.069	.073	.068	.074	.061	.082	.071	.071	.077	.08	.073	.065	.075	.072	.082	.074	.081	
		0.15	.177	.18	.18	.185	.18	.184	.178	.185	.18	.182	.184	.181	.181	.181	.178	.184	.185	.184	.188	
3	3	0.05	.079	.082	.081	.082	.086	.092	.087	.08	.09	.083	.079	.089	.08	.085	.084	.089	.086	.082	.088	
		0.1	.037	.038	.037	.039	.039	.039	.037	.035	.038	.039	.034	.039	.038	.037	.036	.038	.038	.038	.038	
		0.15	.044	.043	.043	.046	.046	.044	.043	.045	.044	.045	.044	.046	.048	.046	.044	.044	.046	.046	.05	
1	1	0.05	.199	.201	.2	.2	.199	.201	.2	.201	.201	.201	.201	.19	.201	.2	.201	.2	.2	.2	.201	
		0.1	.214	.219	.219	.218	.219	.218	.217	.216	.216	.218	.218	.217	.217	.217	.219	.217	.218	.217	.219	
		0.15	.213	.218	.216	.217	.218	.219	.218	.219	.217	.218	.218	.218	.218	.217	.218	.219	.218	.218	.217	
3	2	0.05	.238	.239	.239	.24	.24	.24	.239	.24	.24	.24	.24	.241	.239	.239	.238	.239	.24	.239	.24	
		0.1	.237	.239	.238	.239	.239	.239	.239	.24	.239	.24	.239	.239	.239	.238	.239	.239	.24	.239	.24	
		0.15	.236	.237	.238	.238	.239	.238	.239	.238	.238	.238	.239	.238	.238	.237	.237	.238	.238	.238	.238	
3	3	0.05	.001	.001	.0	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	
		0.1	.232	.231	.23	.233	.232	.233	.235	.233	.234	.233	.233	.232	.233	.233	.233	.235	.234	.233	.234	
		0.15	.223	.226	.223	.228	.227	.227	.226	.227	.23	.227	.227	.227	.227	.228	.228	.229	.226	.227	.228	
1	1	0.05	.056	.055	.056	.055	.056	.056	.056	.055	.055	.056	.056	.056	.056	.056	.056	.056	.056	.056	.056	
		0.1	.176	.178	.178	.178	.178	.178	.178	.176	.178	.178	.178	.178	.178	.178	.178	.178	.178	.178	.178	
		0.15	.211	.21	.213	.213	.213	.213	.213	.213	.212	.212	.213	.213	.213	.213	.213	.213	.213	.213	.213	
4	2	0.05	.033	.034	.035	.034	.035	.034	.035	.035	.035	.035	.035	.035	.035	.035	.035	.035	.035	.035	.035	
		0.1	.075	.076	.076	.076	.076	.076	.076	.075	.075	.076	.076	.076	.076	.076	.076	.076	.076	.076	.076	
		0.15	.218	.216	.218	.218	.218	.218	.218	.218	.218	.218	.218	.218	.218	.218	.218	.218	.218	.218	.218	
3	3	0.05	.064	.064	.064	.065	.065	.064	.065	.065	.065	.065	.065	.065	.065	.065	.065	.064	.065	.065	.065	
		0.1	.049	.05	.049	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.049	.05	.05	.05	
		0.15	.016	.015	.015	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	
d	τ	δ_{\leq}	$d(W^r, W^{r+1}, S)$																			
			24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	
1	1	0.05	.066	.067	.064	.064	.064	.068	.067	.064	.063	.065	.066	.067	.062	.068	.066	.067	.066	.067	.067	
		0.1	.116	.11	.103	.103	.115	.115	.108	.11	.107	.115	.109	.107	.114	.117	.107	.116	.112	.115	.114	
		0.15	.135	.135	.139	.137	.139	.136	.132	.143	.133	.143	.132	.134	.129	.137	.137	.142	.13	.141	.14	
2	2	0.05	.111	.115	.113	.114	.111	.11	.11	.112	.108	.112	.11	.111	.112	.112	.115	.114	.11	.113	.109	
		0.1	.078	.079	.073	.077	.077	.067	.074	.08	.068	.072	.065	.076	.072	.074	.076	.073	.068	.073	.067	
		0.15	.184	.185	.182	.182	.182	.182	.178	.183	.172	.184	.175	.179	.183	.177	.184	.186	.176	.187	.18	
3	3	0.05	.075	.087	.08	.086	.09	.085	.091	.086	.089	.083	.091	.076	.086	.088	.082	.091	.087	.088	.092	
		0.1	.039	.039	.036	.036	.039	.037	.039	.038	.037	.036	.036	.037	.036	.038	.035	.038	.037	.035	.039	
		0.15	.046	.044	.045	.045	.048	.048	.048	.046	.045	.043	.047	.042	.045	.049	.048	.048	.049	.044	.048	
1	1	0.05	.201	.2	.2	.2	.201	.198	.199	.2	.198	.201	.191	.2	.19	.2	.199	.2	.2	.2	.198	
		0.1	.218	.218	.217	.217	.217	.218	.217	.217	.216	.217	.218	.217	.218	.217	.216	.217	.215	.219	.217	
		0.15	.218	.218	.218	.218	.219	.217	.217	.218	.217	.216	.216	.217	.219	.217	.218	.216	.218	.219	.216	
3	2	0.05	.24	.239	.24	.239	.239	.239	.238	.239	.237	.239	.239	.237	.24	.238	.238	.24	.239	.24	.239	
		0.1	.239	.239	.239	.239	.239	.239	.238	.239	.238	.237	.238	.238	.24	.237	.238	.239	.239	.239	.239	
		0.15	.238	.237	.238	.237	.237	.238	.236	.238	.236	.237	.237	.237	.237	.236	.237	.238	.236	.238	.236	
3	3	0.05	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	
		0.1	.233	.234	.232	.235	.233	.232	.233	.231	.233	.233	.233	.231	.232	.232	.234	.235	.233	.232	.235	
		0.15	.229	.229	.228	.228	.228	.227	.228	.227	.226	.228	.226	.227	.226	.225	.228	.228	.228	.228	.226	
1	1	0.05	.056	.055	.056	.056	.055	.055	.055	.055	.055	.055	.054	.055	.056	.056	.056	.056	.056	.056	.056	
		0.1	.178	.178	.178	.178	.178	.178	.178	.178	.178	.178	.178	.178	.178	.178	.178	.178	.178	.178	.178	
		0.15	.213	.213	.213	.213	.213	.213	.213	.213	.213	.213	.213	.213	.213	.213	.213	.213	.213	.213	.213	
4	2	0.05	.035	.035	.035	.034	.035	.034	.035	.035	.035	.034	.035	.034	.035	.034	.035	.035	.035	.035	.035	
		0.1	.075	.076	.076	.076	.076	.076	.076	.075	.076	.074	.076	.075	.076	.076	.076	.076	.076	.076	.075	
		0.15	.218	.218	.218	.218	.218	.216	.218	.218	.218	.218	.218	.218	.216	.218	.218	.218	.218	.218	.218	
3	3	0.05	.065	.065	.063	.065	.065	.065	.065	.065	.065	.065	.065	.065	.064	.064	.065	.065	.064	.065	.065	
		0.1	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	
		0.15	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	.016	

Online Support for Tertiary Mathematics Students in a Blended Learning Environment

Mary Ruth Freislich^{*1}, Alan Bowen-James²

¹ *School of Mathematics and Statistics, University of New South Wales, Sydney, 2052, Australia*

² *Le Cordon Bleu Business School, Sydney, 2112, Australia*

ARTICLE INFO

Article history:

Received: 14 November, 2021

Accepted: 20 February, 2022

Online: 18 March, 2022

Keywords:

Blended learning environments

Scaffolding

SOLO taxonomy

Tertiary mathematics

ABSTRACT

The context for the study was a naturally occurring quasi-experiment in the core mathematics program in a large Australian university. Delivery of teaching was changed in a sequence of two initial core mathematics subjects taken by engineering and science students. The change replaced one of two face-to-face tutorial classes per week by an online tutorial. Tasks in the online tutorial were designed to lead the students through the week's topics, using initially simpler tasks as scaffolding for more complex tasks. This was the only change: syllabus and written materials were the same, as was students' access to help from staff and discussion with peers. The study compared learning outcomes among students in two adjacent years: Cohort 1, the last before the change, and Cohort 2, in the first implementation of the change to a blended learning environment. Learning outcomes were assessed by a method derived from the SOLO taxonomy, which used a common scale for scoring written answers to examination questions in the two cohorts. In the first mathematics subject students doing online tutorials had significantly higher scores than those studying before the change. In the second mathematics subject there were no significant differences. The conclusion was that the online tutorials gave an advantage to students beginning university study and gave adequate support to those in the subject taken a little later. It can be concluded that the use of an online teaching component in the delivery of university mathematics programs is not only justifiable but desirable, subject to careful design of the teaching material offered.

1. Introduction

This paper is an extension of work originally presented at the *IEEE 2020 International Conference on Computer Science and Computational Intelligence* [1]. Extended content is mainly in the research background section, dealing with scaffolding in blended learning environments and the importance of the design of online teaching material. It includes discussion of the unsuitability for mathematics learning of existing instruments used to evaluate students' approaches to studying. Results contain effect sizes. The discussion includes more detail, and a conclusion section has been added.

The reported project stems from a change in first year mathematics teaching in a large Australian university, which replaced one out of two face-to-face tutorials with an online tutorial consisting of a set of tasks designed to lead the student

through the required material step by step from simpler to more complex tasks. The change was made in the core sequence of first year mathematics subjects, Mathematics 1A and Mathematics 1B, which is taken by science and engineering students. Before the change, teaching was entirely face-to-face, consisting of lectures to large groups, plus two tutorials given to problem solving and answering students' questions, in much smaller groups.

The only change in the organization of teaching was the replacement of one face-to-face tutorial by an online tutorial. In the online tutorial, immediate feedback identified errors without giving solutions. Completing the tutorial tasks earned a small contribution to the student's final mark.

The study was carried out before higher education was disrupted by the Covid19 epidemic, and apart from the change in one tutorial, there was no other change in the organization and material in the core sequence. That is, the syllabus and the written

^{*}Corresponding Author: Mary Ruth Freislich, m.freislich@unsw.edu.au

teaching material did not change, and the online tasks were very similar to those used in face-to-face tutorials. All students had access to help from staff and discussion with peers was still available in the face-to-face teaching groups. The study dealt with two cohorts of students in adjacent years, the last year before the change, and the first year of its introduction. Admission criteria had not changed, and the secondary school mathematics course taken by local students had not changed. It is therefore reasonable to conclude that the two cohorts were comparable in background and level of selection. In the year of the change, survey results indicated that students were satisfied with the teaching delivery.

It seems, therefore, that the natural quasi experiment afforded by the change offered good control for a study of any effect that the change might have on the quality of learning outcomes among students in the two cohorts. Making a comparison requires a valid method for evaluating outcomes, and validation requires the use of observable evidence. The course evaluation procedures advocated in improvement programs by the Accreditation Board for Engineering and Technology [2] emphasize the importance of direct assessments.

The present study uses observable evidence from students' final examination scripts. The comparison is based on a method described below in Section 2. It is emphasized here that the method has potential importance because it is direct and criterion-based. Before giving detail about the method, one needs to examine existing research that suggests directional predictions that can be tested.

2. Background

2.1. Students' use of resources

The newly implemented program studied in the present work represents provision of an online learning resource, rather than online instruction, given that the majority of instruction was face-to-face, and the online segment involved students' work rather than online instruction. This means that the program is not comparable with projects involve online instruction such as that reported in [3].

In [4] the authors make the important point that evaluation of any new learning resource can be invalidated if there is no evidence about whether students have used the resource. Findings in [5] were that mathematics students who were offered several optional learning resources, they tended to use only one of the set of resources. It is worth clarifying that the present study deals with only one new resource, so that the problem of choice does not arise. There is also no uncertainty about whether the new resource was used because students' work online leaves an audit trail.

2.2. Student learning and the transition to university mathematics

There are three research strands that are important to the purpose of the present work.

First, in the review of research on student learning made in [6], it is noted continuing importance is found for an approach to learning that contains a continuing purpose of understanding material and attempting to link and compare different ideas. Understanding and high quality of learning are unlikely when students' approach is atomistic, focused on the accumulation of unrelated detail. For mathematics students, understanding is achieved and tested by active problem-solving. The instruments used in the British and Australian research reviewed in [6] measure the search for understanding using items describing wide reading and venturing beyond the syllabus. Among undergraduate mathematics students, such items are irrelevant for all but the most highly gifted students, but at all levels of talent the intention and achievement of understanding relate to activity in doing mathematical tasks appropriate to the level of study.

Active problem solving requires effort and persistence, which relates to North American research underlying the *National Survey of Student Engagement (NSSE)*, [7]. This research strand found that self-regulation of study is very important to students' learning, and this is an obvious requirement for activity in working on mathematical tasks.

Directly relevant Australian work on student engagement in mathematics learning is supported by the theoretical outline given in [8]. The authors note that student engagement in mathematics is multidimensional, with fuzzy boundaries between categories. The principal components that have been identified as affecting student engagement are consistent with components of the student learning research. *Expectancy-value theory* is defined by the value given to a learning goal and the learner's expectation of achieving it. This entails interest factors and confidence, as well as practical reasons for valuing achievement, and research has shown that it relates to quality of learning outcome. Similar importance is attached to the factor of self-regulated study identified in the North American research.

University students' choice of mathematics requires some previous success, and the choice implies that they attach some value to the subject. But there is no lack of evidence that many students find the transition to university mathematics very difficult, and the evidence comes from a wide variety of settings. Examples are afforded by the work reported in [9], for Britain, in [10] for Sweden and in [11] for Australia. Beginning university students can find self-regulation difficult. For mathematics students, previous levels of motivation, goal setting and self-regulation will not be sustained at university level if successful mathematical activity is not sustained.

Experience of difficulty may lead to discouraged and anxious avoidance of attempted engagement with mathematical tasks. The work reported in [12], done in an Australian setting indicates, that beginning mathematics students can benefit from learning support

that facilitates engagement with mathematical tasks. The next requirement is for evidence relevant to beneficial types of support.

2.3. Scaffolding and transfer of responsibility

The material for the online tutorials was specially designed to lead the student through the week's mathematical topic using a sequence of tasks that progressed from simple to more complex. Immediate feedback was given for each response, informing the student only whether the response was correct or not, without giving a solution. In addition, the sequences of tasks were designed so that solutions to earlier tasks could help with the later more difficult tasks. The online work could be done in multiple sessions within a specified time period, so the student could temporarily leave a task to look up material, ask for help, or discuss it with peers. Such a design has the potential to function as *scaffolding* for the extension of students' understanding.

Scaffolding is defined as intervention by a teacher to support students in achieving a learning goal that they would be unlikely to achieve without support [13]. There has been considerable discussion of the method of intervention and the design of the teacher's intervention, so that it extends the student's own reasoning without imposing or supplying a solution. The original idea rests on Vygotsky's thesis, described in [14], that the most valuable instruction is that which leads a learner into a development defined as being in the *zone of proximal development*. That is, the learner is already on the border of extended capability, and hence can reach extension with minimal appropriate help. The idea of scaffolding is defined by interaction between teacher and student. The authors in [15] found that interactive scaffolding led by the teacher in relatively small community college mathematics classes was very much more successful than previously used approaches. For large enrolment groups, limits of resourcing make the original form of scaffolding impossible. But it is argued here that the design of the online tutorials affords an approximation to scaffolding, because the gradient of task difficulty and the immediate feedback provide indirect assistance in the extension of understanding, with the limitation of the feedback also implying that assistance is not too intrusive.

Transfer of responsibility to the learner is also an important underlying goal of providing scaffolding [16]. It is pointed out in [17] that, for scaffolding to make its widest contribution, it needs a definition that empowers the learner, so that the student becomes independent of the presence of an insightful teacher as agent. In the context of mathematics, they propose problem solving as the means of creating self-scaffolding. In contrast to face-to-face tutorials, online tutorials give all responsibility for work on the given tasks to the student, with the minimal assistance designed to foster effort and persistence. Organizational responsibility in scheduling time is also required, but the important factor is the design of the tasks facilitating active engagement in the tasks,

which serves to build the understanding, independence and self-regulated study found important in the studies described here and in Section 2B.

2.4. Blended mathematics teaching and the importance of design

Evidence is available that well-designed online materials can function in this way. Studies of statistics programs [18], [19] indicate that achievement gains follow careful adjustment of materials, designed to integrate the learning environment consistently, and to foster understanding. The results of [18] are particularly important, because the material provided to students was revised from year to year, and benefits to students' achievement appeared only in later years. These results are compatible with the established distinction between medium as a means of delivery and the designed study program as the goods delivered [20] Rapid and flexible delivery can give an advantage only if goods of value are delivered.

The study described in [21] is also highly relevant to the idea of scaffolding afforded by suitably designed material. It deals with a very large group of statistics students of variable academic and national background, who were offered an online tutorial system that proceeded from diagnostic testing to select tasks best adapted to each student's stage of learning. The study found that the time spent using the online program was positively related to achievement, with the strongest effect among students whose scores on Vermunt's *Inventory of Learning Styles* [22] indicated that they were less well adapted to university study.

2.5. Assessing learning outcomes

In Australian work on learning outcomes, [23] the researchers developed a classification of the quality of learning outcomes based on actual responses to a variety of educational tasks. The classification used criteria defined by the complexity, adequacy of coverage, and consistency of observable responses to set tasks. They defined a system of levels of outcome called the *Structure of Observed Learning Outcomes (SOLO) Taxonomy*. The value of the reference to the observable is clear. The researchers claimed that the classification was invariant across disciplines and justified the claim by giving illustrations from the work done in the principal areas of school study, across the middle years of schooling, from upper primary level to junior secondary. The SOLO levels, as defined in [23] are listed in Table 1 below.

The SOLO split between the *Multistructural* and *Relational* levels is based on consistency in reasoning, and so reflects the dichotomy between understanding relationships and atomistic display of facts which is of obvious importance in mathematics, with achievement of the relational level providing evidence of understanding. The wide applicability of the SOLO taxonomy is not relevant to the present study, but the issue of consistent reasoning is central to it. The applicability of SOLO to

mathematics was based on research that identified patterns of errors and misconceptions in students' mathematics learning.

Table 1: The SOLO taxonomy

Level	Definition
Prestructural	No valid response
Unistructural	One aspect of the problem correctly identified, but no diversity of aspects presented, so that questions of consistency cannot arise.
Multistructural	Multiple relevant information presented and used, but without considering relationships between different parts, so that inconsistency appears.
Relational	Multiple relevant information presented and used in a way that recognizes relationships and achieves consistency within the given task.
Extended abstract	Multiplicity recognized and consistency achieved over a context beyond that of the given task.

The SOLO taxonomy has been used at tertiary level as a framework for defining intended learning outcomes for programs in mathematics and computer science [24] and its application in other science disciplines at tertiary level has been found to be a valuable diagnostic tool [25]. The SOLO levels were adapted for the work reported in [26] to define a method of evaluating levels of learning outcomes in tertiary students' mathematics.

The focus was on examination performance in early undergraduate years, so the highest SOLO level was not considered relevant. The other four levels were used to construct a scoring system intended to provide a common scale usable across tasks involving the same mathematical material, examined at a similar level of difficulty.

The criteria used were, first, logical consistency, and second, adequate coverage of the task. A student's response to an examination question was assigned to one of six levels, labelled from 0 to 5. Levels 4 and 5 required the logical consistency of the SOLO relational level, with 5 given for a completely correct solution, and 4 given if there was a small error that did not affect consistency, like a minor slip in arithmetic or a copying error. Levels 0 and 1 correspond to SOLO Prestructural and Unistructural levels: nothing right or only one relevant aspect of the problem identified. Solutions with an error of logic at the Multistructural SOLO level, with more than one good step presented, were classified as level 2 or 3, depending on how much of a satisfactory solution was present. Examination questions were split into self-contained tasks, and each was scored independently. A composite score was obtained by summing the task scores, weighted using the proportion of the examination marks assigned to each.

The method does not attempt the generality claimed for the SOLO taxonomy. Validity is claimed only for the close relationship between tasks, depending on the stability of syllabus, staffing, student intake, teaching materials and most of the implementation of teaching in the two adjacent year groups. The SOLO taxonomy is well adapted to mathematics because its criteria fit the requirements of mathematical tasks. But its most important characteristic is its being defined in terms of the observable. The North American Accreditation Board for Engineering and Technology [2]) argues that a teaching program cannot be adequately evaluated without a direct method for examining students' learning outcomes, one which is closely fitted to the actual study program, both of which requirements apply to the method described. Applying the scoring method is similar to examination marking, and scores correlate at over 0.9 with examination marks, which implies similar ranking. What the method is intended to achieve is a common ranking for the two year-groups' performance on similar tasks. It is worth noting also that a direct method of examining learning outcomes has advantages over the use of questionnaires to assess approaches to studying. Two reasons are important. The first is intrinsic: direct assessment avoids problems associated with the reliability of self-reported data about behaviour and attitudes. The second reason is the mismatch between the existing instruments used to assess approaches and the study of mathematics, at least at undergraduate level. This has already been mentioned in connection with the approach instruments described in [6]. But one should also note that similar remarks apply to the North American *NSSE*, and the Australian Survey of Student Engagement (*AUSSE*) derived from it [27] derived from it.

The point here is that the *AUSSE* measure higher level thinking by items dealing with extended essay- style writing and multiple revision of drafts. In the development work for the *AUSSE*, it was found [27] that science students had low scores of higher-level thinking, but it is probable that such results are contaminated by the inadequacy of the instrument.

3. Method

3.1. Sample

The target population was the set of students enrolled for Mathematics 1A and 1B, in adjacent years, taking the groups from the first time in each year that the unit was offered. Simple random samples were drawn from those students who sat the final examination. This means that those who did not survive to the final examination could not be considered, but this restriction applies to all the groups being compared. Questions involving students' gender were not part of the study, but gender information was available, and was recorded, because any gender-related patterns that might emerge would be of interest. Sample numbers are in Table 1. The proportions of females and males in the sample are very similar to proportions in the total groups.

Table 2: Sample

Cohort	Mathematics 1A		Mathematics 1B	
	Female	Male	Female	Male
1 (from the last year before the change)	53	152	38	142
2 (from the first year when the change was introduced)	49	154	44	153

3.2. Analyses

The two cohorts were compared within each of the two mathematics subjects. In each subject, four groups defined by cohort and gender were compared using analysis of variance. In the case of a significant overall result, differences between groups were examined using least significant differences. For cases where there were significant results, effect sizes were calculated. Analyses were done using the open-source package Rstudio [28].

4. Results

4.1. Mathematics 1A

Descriptive statistics are in Table 3, and the analysis of variance data are in Table 4.

Table 3: Mathematics 1A Descriptive statistics

Cohort		Female	Male
1 All teaching face-to-face	Mean	10.66	10.07
	St. dev.	3.58	3.62
	<i>n</i>	63	152
2 Blended teaching	Mean	12.87	12.24
	St. dev.	3.93	3.73
	<i>n</i>	49	154

Table 4: Mathematics 1A Analysis of variance

Analysis of variance				
Source	Sums of squares	df	Mean squares	F
Between groups	521.29	3	173.76	12.74***
Residual	5510.47	404	13.64	
Total	6031.44	407		

*** $p < 0.001$

The means for Cohort 2 are higher than those for Cohort 1, and the analysis of variance gives a high level of significance to differences between groups. Results for comparisons between pairs of groups using least significant differences are in Table 5

Table 5: Least significant differences

Groups in order of means			
Cohort	1 female	2 male	2 female
1 male	0.90	5.14***	4.62***
1 female	t	2.79**	3.10**
2 male			1.04

** $p < 0.01$; *** $p < 0.001$

The differences are significant for all comparisons of Cohort 1 groups with Cohort 2 groups, and no within-cohort comparisons between females and males were significant. The purpose of the study did not include gender comparisons but grouping by gender was in the analysis because it was possible that different delivery of teaching might have different effects for females and males.

Effect sizes for the four significant comparisons are in Table 6. The interpretations use the classification described in [29] and are high or very high in all cases.

Table 6: Effect sizes

Group	Cohort 2 male		Cohort 2 female	
	Effect size		Effect size	
Cohort 1 male	0.60	Very high	0.77	Very high
Cohort 1 female	0.46	High	0.46	High

4.2. Mathematics 1B

Descriptive statistics are in Table 7 and the analysis of variance results are in Table 8. There were no significant differences between groups in Mathematics 1B.

Table 7: Mathematics 1B: Descriptive statistics

Cohort		Female	Male
1 All teaching face-to-face	Mean	11.48	11.46
	St. dev.	3.77	3.74
	<i>n</i>	38	142
2 Blended teaching	Mean	11.82	11.32
	St. dev.	3.46	3.44
	<i>n</i>	44	154

Table 8: Mathematics 1B Analysis of variance

Analysis of variance				
Source	Sums of squares	df	Mean squares	F
Between groups	8.52	3	2.84	0.82 ns
Residual	4819.82	373	12.92	
Total	4828.34	378		

5. Discussion

It is worth noting here again that no gender differences were found. Marginally higher mean scores for females probably only reflect the higher selection of the female groups, given that tertiary mathematics groups still contain considerably more males.

It is clear that the results for Mathematics 1A show advantages in the online component of delivery of teaching. The advantage is in the direction predicted from the research background, subject to the importance of the design of the online teaching material. The digital audit trail afforded by the technology gives assurance that the online learning resource was used by the students, which functions as an additional control factor. Because Mathematic 1A is the first core mathematics subject taken by engineering and science students, one can

conclude that the online program facilitated students' transition to university study.

The lack of significant differences between the two cohorts in Mathematics 1B can be explained by combining evidence from the literature with the conclusion given for Mathematics 1A. Mathematics 1B is the second subject in first-year core mathematics, its students are at least one semester further into university study than most students in Mathematics 1A and are more highly selected because they have already passed Mathematics 1A. In [21] the findings indicated that online resources were most helpful to students who were initially less well adapted to university study. The mathematics 1B groups, therefore, are likely to have less need of help than students who are mostly new to university study.

But the finding for Mathematics 1B is still useful evidence because it indicates that the online program shows no disadvantage compared with fully face-to face teaching. This means that, if one regards the medium as a delivery vehicle, the results indicate delivery of adequate goods. The speed and flexibility of delivery therefore become relevant. The audit trail permitted by the technology also enables improvement of the online material through tracking areas where students have most difficulty. The method of comparison of outcomes used in the present study can be used to compare different sets of online material, serving as the direct Students' written assessments can be scanned into digital records, which opens the way to a cyclic use of technology to provide research material for evaluation of what the technology delivers. Such material would also permit research on changes in students' learning over some years.

In a review of research on fully online teaching of undergraduate mathematics, [30] it is reported that the results are mostly unfavorable to online teaching. It should be emphasized, however, that the present study represents a different field, because the blended learning environment involved retained easy contact with staff and peers. That is to say, the learning environment was not exclusively online and, indeed, assumed a degree of offline interpersonal engagement.

It should also be noted that the direct assessment of quality of learning outcomes in the present study has advantages over alternative methods. It clearly is unreasonable to judge online teaching using correlations between results of assessments of different teaching components, but the use of grades alone also does not provide a clear determination of efficacy. Hence, results of the study [31], which used grades, cannot be considered as corroboration for the present study.

It was noted In Section 2B that instruments used to assess students' approaches to studying are not well adapted to mathematics learning. The underlying concept of the value of a search for understanding is clearly important in all fields, so that, even after some decades of stabilization of existing instruments,

adaptation to mathematics would be useful. Records from online tutorial tasks and written examinations could be combined with initial qualitative investigation of students' approaches to and experience of studying mathematics.

6. Conclusion

The results indicate that the use of an online component in the delivery of first year tertiary mathematics can be justified as producing enhanced learning outcomes among beginning students, and no disadvantage to those at a slightly later stage, provided that the online teaching material is carefully designed to lead the students from simpler to more complex tasks. Hence any recommendation for the extended use of online teaching material, and any future research of online mathematics teaching, must focus primarily on the quality of the design of that material.

The present study is limited to first-year mathematics. It follows that investigation in other contexts and later stages of university study would be a necessary supplement. The increasing use of online teaching delivery affords the opportunity for such work. In addition, one should note that the technology furnishes detailed records of students' use of materials and performance on assessment tasks that provide valuable data for study.

The present study did not address students' experience of studying. In the background section it was noted that existing self-report questionnaires on students' approaches to studying are unsuitable for mathematics learning. The development of suitable instruments with a similar purpose, but targeting more appropriate approaches, is an open field. The development of such instruments would be facilitated by initial exploratory work using qualitative methods to elucidate salient aspects of students' experience of studying mathematics.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors wish to acknowledge the support of the University of New South Wales and Le Cordon Bleu Australia.

References

- [1] M.R. Freislich, A. Bowen-James, "Observable learning outcomes among tertiary mathematics students in a newly implemented blended learning environment". In 2020 IEEE International Conference on Computer Science and Computational Intelligence (CSCI), 976-980, 2020, doi:10.1109/CSCI151800.2020.00181.
- [2] Accreditation Board for Engineering and Technology, 2020, *ABET accreditation*. <https://www.abet.org>.
- [3] S. Lambert, "Reluctant mathematicians: skills-based MOOC scaffolds a wide range of learners", *Journal of Interactive Media in Education*, **2015**(1), 1-11, 2015 doi.org/w5334/jime.bb.
- [4] P. Sharma, M.J. Hannafin 2007, "Scaffolding in technology enhanced learning environments", *Interactive Learning Environments*, **15**(1), 27-46, 2007, doi:10.1080/10494820600996972.
- [5] M. Inglis, A. Palipura, S. Trenholm, J. Ward, 2011, "Individual differences in students' uses of optional learning resources", *Journal of Computer Assisted Learning*, **27**(6), 490-502, doi:10.1111/j.1365-2729.2011.00417.x.

- [6] J.T.E.Richardson, 'Student learning in higher education: A commentary', *Educational Psychology Review*, vol. 29, pp. 353-362, 2017, doi: 1007/s10648-017-9410-x.
- [7] Indiana University School of Education, National Survey of Student Engagement, University of Indiana, 2017.
- [8] H.M. Watt, M. Goos, M, "Theoretical foundations of engagement in mathematics" *Mathematics Education Research Journal*, **29**(2), 133-142, 2017, doi:10.1007/s13394-017-0206-6.
- [9] M. McAlinden, A. Noyes, "Mathematics in the disciplines at the transition to university", *Teaching Mathematics and its Applications*, **38**(2), 61-73, 2019, doi.org/10.1093/teamat/hry004.
- [10] S.H. Bengmark, H. Thunberg, T.M. Winberg, 2017, "Success-factors in transition to university mathematics", *International Journal of Mathematical Education in Science and Technology*, **48**(7), 988-1001, 2017,doi: 10.1080/0020739X.2017.1310311.
- [11] P.W. Hillock PW, R.N. Khan, 2019, "A support learning programme for first-year mathematics", *International Journal of Mathematical Education in Science and Technology*, **50**(7), 1073-1086, 2019, doi:10.1080/0020739X.2019.16569026.
- [12] L.J. Rylands, D. Shearman, "Mathematics learning support and engagement in first year engineering", *International Journal of Mathematical Education in Science and Technology*, **49**(8), 1133-1147, 2017, doi: 10.1080/0020739X.2018.1447699.
- [13] C. Quintana, 2021, 'Scaffolding inquiry', in R.G. Duncan, & C.A. Chinn, (Eds.), *International handbook of inquiry and learning*, Routledge, 176-188.
- [14] B.H. Johnsen, BH, 2020, 'Vygotsky's Legacy Regarding Teaching-Learning Interaction and Development', In B.H. Johnsen (ed.), *Theory and Methodology in International Comparative Classroom Studies*, Cappelen Damm Akademisk, 82-98.
- [15] T. Gula, C. Hoesler, W. Majciejewski, W, "Seeking mathematics success for college students: A randomised field trial of an adapted approach", *International Journal of Mathematical Education in Science and Technology*, **48**(8),127-143, 2021, doi: 10.1080/0020739X2015.1029026.
- [16] B.R. Belland, A.E. Walker, M.W. Olsen H. Leary, H., "A pilot meta-analysis of computer-based scaffolding in STEM education", *Educational Technology and Society*, **18**(1), 183-197, 2015, <https://www.jstor.org/stable/jedtechsoci.18.1.18>.
- [17] D. Holton, D. Clarke, D., "Scaffolding and metacognition"; *International Journal of Mathematical Education in Science and Technology*, **37**(2), 127-143, 2006, DOI:10.1080/0020730500285818.
- [18] L. Zetterqvist, "Applied problems and use of technology in an aligned way in basic courses in probability and statistics for engineering students: a way to enhance understanding and increase motivation," *Teaching Mathematics and its Applications*, **36**(2), 108-122, 2017, doi.org/ 10.1093/teamat/ hrx004
- [19] A.H. Jonsdottir, A.A. Bjornsdottir, G. Stefansson, G., "Difference in learning among students doing pen-and-paper homework compared to web-based homework in an introductory statistics course", *Journal of Statistics Education*, **25**(1), 12-20, 2017, doi: 10.1080/10691898/2017.1291289.
- [20] R.E. Mayer, "Thirty years of research on online learning", *Applied Cognitive Psychology*, **33**(2), 152-159, 2019, doi: 10.1002/acp.3482.
- [21] D.J. Tempelaar, B. Rienties, B. Giesbers, "Who profits most from blended learning?" *Industry and Higher Education*, **23**(4), 285-292, 2009, doi.org/10.1145/3170358.3170385.
- [22] J.D. Vermunt, Y.J. Vermetten, YJ, 2004, 'Patterns in student learning: relationships between learning strategies, conceptions of learning and learning outcomes', *Educational Psychology Review*, **16**(4), 359-384, 2004, doi:org/10.1007/s10648-004-0005-y.
- [23] J.B. Biggs, K.F. Collis, KF (2014). *Evaluating the quality of learning: The SOLO taxonomy*. Academic Press, 2014.
- [24] C. Braband, B. Dahl, B, "Using the Solo taxonomy to analyze competence progression in science curricula", *Higher Education*, **58**(4),.531-549, 2009, doi: 10.1007/s10734-009-9216-4.
- [25] L.C. Hodges, L. Harvey, "Evaluation of student learning in organic chemistry using the SOLO taxonomy", *Journal of Chemical Education*, **80**(7) 785-787, 2003, doi:org/10.1021/ed080 p785.
- [26] M.R. Freislich, A. Bowen-James, 2019 "Effects of a change to more formative assessment among tertiary mathematics students", *Anziam Journal Electronic Supplement*, **61**, C255-C271, doi:10.21914/anziamj.v6110.15166.
- [27] A. Radloff, H. Coates, 2009, *Doing more for learning: Enhancing engagement and outcomes*. Australian Council for Educational Research, 2009.
- [28] RStudio, Open source and professional software for data science, <<https://rstudio.com>>, 2020
- [29] S. Higgins, M. Katsipatakis, 2016, "Communicating comparative findings from meta-analysis in educational research: some examples and suggestions", *International journal of Research and Method in Education*, **30**(3), 237-254, doi: 10/1080/ 1743727X. 2016.1166486.
- [30] S. Trenholm, J. Peschke, "Teaching undergraduate mathematics fully online: a review from the perspective of communities of practice", *International Journal of Educational Technology in Higher Education*, **17**, Article 37, 2020, doi:org/10.1186/s41239-020-00215-0.
- [31] B. Loch, R. Borland, N. Sukhurovka, 'Implementing blended learning in tertiary mathematics teaching', *The Australian Mathematical Society Gazette*, **46**(2), 90-102, 2019.

Appendix: Examples of the scoring method

A. Algebra 1

Find conditions on b_1, b_2, b_3 to ensure that the following system of equations has a solution.

B. Algebra 2

a) Find all roots in the complex numbers of

$$z^5 + 1 = 0$$

$$\begin{matrix} x + 2y & . & . & = & b_1 \\ x + y - z & = & b_2 \\ 2x + y - 3z & = & b_3 \end{matrix}$$

Solution

$$\begin{pmatrix} 1 & 2 & 0 & b_1 \\ 1 & 2 & -1 & b_2 \\ 2 & 2 & -3 & b_3 \end{pmatrix} \text{ Step 1 } \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 0 & b_1 \\ 0 & -1 & -1 & b_2 - b_1 \\ 0 & -3 & -3 & b_3 - 2b_1 \end{pmatrix} \text{ Step 2 } \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 0 & b_1 \\ 0 & -1 & -1 & b_2 - b_1 \\ 0 & 0 & 0 & b_3 - 3b_2 + b_1 \end{pmatrix}$$

Conclusion

Solutions exist if and only if

$$b_3 - 3b_2 + b_1 = 0$$

Table 9 Scoring examples for Algebra 1

Score	Example
5	All correct
4	Step 1 correct. Step 2; $\begin{pmatrix} 1 & 2 & 0 & b_1 \\ 0 & -1 & -1 & b_2 - b_1 \\ 0 & 0 & 0 & b_3 - 5b_2 + b_1 \end{pmatrix}$ [Mistake in arithmetic.] Conclusion: solutions exist if and only if $b_3 - 5b_2 + b_1 = 0$
3	Row operations correct to the end of Step 2. But conclusion given as: $b_1 \neq 0, b_1 \neq b_2, b_3 - 3b_2 + b_1 \neq 0$
2	Row operations correct to the end of Step 2. No conclusion.

1	Step 1 correct. Then replace Row 2 by Row 2 + (1/2) Row 1, giving $\text{Row 2} = (0 \ 0 \ -1 \ b_2)$ [This shows row operations are not understood.]
---	---

Factorise $z^5 + 1$ over the complex numbers.

Factorise $z^5 + 1$ over the real numbers.

Solution

Put $z = re^{i\theta}$. Then

$$r^5 e^{5i\theta} = 1 e^{(\pi+2k\pi i)}$$

So $r = 1$ and

$$5\theta = \pi + 2k\pi. \quad \theta = \frac{(2k+1)\pi}{5}$$

Distinct solutions occur for

$$k = 0, 1, -1, 2, -2.$$

So the solutions to the equation are:

$$e^{\frac{\pi i}{5}}, \quad e^{\frac{3\pi i}{5}}, \quad e^{-\frac{\pi i}{5}}, \quad e^{\frac{5\pi i}{5}} = -1, \quad e^{-\frac{3\pi i}{5}}$$

$$z^5 + 1 =$$

$$(z + 1) \left(z - e^{\frac{\pi i}{5}} \right) \left(z - e^{-\frac{\pi i}{5}} \right) \left(z - e^{\frac{3\pi i}{5}} \right) \left(z - e^{-\frac{3\pi i}{5}} \right)$$

$$z^5 + 1 =$$

$$(z + 1) \left(z^2 - 2z \cos\left(\frac{\pi}{5}\right) + 1 \right) \left(z^2 - 2z \cos\left(\frac{3\pi}{5}\right) + 1 \right)$$

Table 10 Scoring examples for Algebra 2

Score	Example
5	All correct
4	Correct (a), (b), then (c) $(z + 1)(z^2 - 4z + 1)(z^2 - 2 \cos \frac{3i\pi}{5} z + 1)$
3	Correct (a), (b), then (c) $(z + 1)(z - e^{\frac{i\pi}{5}})(z - e^{-\frac{i\pi}{5}})(z - e^{\frac{3i\pi}{5}})(z - e^{-\frac{3i\pi}{5}}).$
2	Roots given as $e^{\frac{2ik\pi}{5}}$ then (b) corresponding to this, no (c)
1	$z = e^{\frac{2k\pi i}{5}}$ and no more

C. Calculus 1

- a) State the Mean Value Theorem
- b) Use the theorem to prove $\sinh x > x$ for $x > 0$

Solution

a) If $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b) , then there exists c in (a, b) such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

b) $f(x) = \sinh x$ is continuous and differentiable everywhere and $f'(x) = \cosh x$. So, there is $c \in (0, x)$ such that

$$\begin{aligned} \frac{\sinh x - \sinh 0}{x - 0} &= \frac{\sinh x}{x} \\ &= \cosh c = \frac{e^c + e^{-c}}{2} > 1 \end{aligned}$$

It follows that $\sinh x > x$ for $x > 0$

Table 11 Scoring examples for Calculus 1

Score	Example
5	All correct
4	Correct up to $\frac{\sinh x}{x} = \cosh c$, then a sketch showing $\cosh x > 1$, but conclusion stated as $\frac{\sinh x}{x} > 0$, and hence $\sinh x > x$
3	Correct up to $\frac{\sinh x}{x} = \cosh c$, then “ $\cosh c < 1$ so $x < \sinh x$ as required.”
2	Correct statement of the theorem, no more
1	Ratio formula for the theorem stated, no conditions, no more

D. Calculus2

Determine whether the following improper integral converges. (Give reasons for your answer.)

$$\int_1^{\infty} \frac{1}{\sqrt{1+x^6}} dx$$

Solution

$$\frac{1}{\sqrt{1+x^6}} < \frac{1}{\sqrt{x^6}} = \frac{1}{x^3}$$

$$\int_1^{\infty} \frac{1}{x^3} dx = \lim_{R \rightarrow \infty} \int_1^R \frac{1}{x^3} dx.$$

$$= \lim_{R \rightarrow \infty} \left(-\frac{1}{2R^2} + \frac{1}{2} \right) = \frac{1}{2}$$

So, the original integral converges, by the comparison test.

Table 12 Scoring examples for Calculus 2

Score	Example
5	All correct
4	Chosen comparison right, but integral evaluated as $-1/4x^2$, with consistent valid conclusion
3	Evaluation of $\int_1^{\infty} \frac{1}{x^3} dx$ correct, but no comparison made.
2	Wrote $\frac{1}{\sqrt{1+x^6}} < \frac{1}{x^3}$ but no more
1	Stated $\int_1^{\infty} f(x) dx = \lim_{R \rightarrow \infty} \int_1^R f(x) dx$ but no more

Solar Energy Assessment, Estimation, and Modelling using Climate Data and Local Environmental Conditions

Clement Matasane^{1,*}, Mohamed Tariq Kahn²

¹Cape Peninsula University of Technology (CPUT), Electrical and Electronic Engineering Department, Symphony Way, Bellville Campus, Bellville, 7925, South Africa

²Cape Peninsula University of Technology (CPUT), Research Chair: Energy, Director: Energy Institute, Head: Centre for Distributed Power and Electronic Systems, Head: Centre for Research in Power Systems, Symphony Way, Bellville Campus, Bellville, 7925, South Africa

ARTICLE INFO

Article history:

Received: 17 July, 2021

Accepted: 11 February, 2022

Online: 18 March, 2022

Keywords:

Solar Energy

Radiation

Insolation

Climate Data

Potential energy

Geographical Parameters

ABSTRACT

On Renewable Energy (RE), this field covers the most significant share of the world energy demand and challenges on the expensive measurement and maintenance equipment to be used. In all studies and designs, global solar radiation (GSR) measurements require assessment, estimation, and models to be applied together with the environment and meteorological data on installing stations at the specific location. These meteorology stations provide measured data throughout the year/ annually or at specified periods, depending on the site of interest. This study includes assessment and estimations of the solar radiation at the Vhembe District using the geographical data measured daily, monthly, and throughout a year in the area. It provides variables such as the geographical maps of the solar availability at a minimum and maximum temperatures obtained during the annual analyses. Determining the solar radiation at a specific location for energy generation involves several procedures, estimations, and calculations using the climatological weather data measurements through MATLAB simulations. In addition, the Geographical Remote Sensing (RS) and Mappings, and Spreadsheet Graph Analytics, were applied to the measured data from the nine installed Weather Stations (WS) in the Vhembe District area was used. The analysis determines the minimum and maximum solar radiation equations associated with the local climate patterns in accommodating the theoretical bases and period changes. The paper contributes to the main project objectives on renewable energy assessment for potentials and generation at a micro/small scale in the district. These parameters are fundamental in estimating and determining the potential solar energy radiation using its extraterrestrial solar radiation per day/ weekly/ monthly. Annual periods towards methods to develop micro/small energy projects for rural and urban communities for domestic and commercial use. As a result, the meteorology analysis is being presented in this study.

1. Introduction

This paper is an extension of work initially presented at the 2019 IEEE PES/IAS Power Africa Conference held in Abuja, Nigeria [1]. This article provides an extension and detailed result to determine the daily, monthly, annual, and solar potential and radiation within the Vhembe District. This demonstrates the

estimating of the energy potential as part of the sub-energy potentials obtained from the wind, biomass/biogas, and hydro energy for the optimal energy generation in the area.

Solar energy applications play a significant role in health, agriculture, civil engineering, and the environment for their execution to support the energy demands within the domain [1], [3]. Hence, evaluating the solar energy potential at any specified location requires accurate solar radiation information. The sunshine duration from the most common variable for predicting

*Clement Matasane, CPUT, Symphony Way, Bellville, 7925, South Africa, +27(021) 4603383 & matasanec@cput.ac.za

global solar radiation (GSR), so sunshine duration can be easily calculated, reliable, and widely used. In increase, solar radiation is the primary root of energy and varies per amount of energy received at different locations [2], [28], [30]. The current developments towards sustainable energy savings and generation using the solar photovoltaic (PV) units have accelerated the maturation process and investment in the area [3]-[5], [11], [29]. In the Limpopo Province (coordinates as 22°50 "22. 08" S and 30°18 "36" E), there is adequate sunlight, which can be more utilized for solar energy applications as shown in Figure 1 to Figure 3. The images were obtained from the Global Solar Atlas developed by the Energy Sector Management Assistance Program (ESMAP) and SOLARIS supported by the World Bank. It is thus essential to harvest and store this natural resource to find a solution to energy shortages and environmental degradation at the state. It is of the view that solar energy systems are considered the most cost-effective and economic power systems in providing off-grid electricity generation in rural areas in the province.

Estimating the renewable energy (RE) potentials using geographical and climatological data requires thorough calculations per specified location. Hence the quantity of solar energy per location is essential, as shown aside from the direct average irradiation, global horizontal irradiation, and potential photovoltaic power for the region from Figure 1 to Figure 3.

More geographical, climatic, and analysis data were added to present conditions and their placement in this paper. Figure 1 to Figure 3 illustrates the available solar map available in the provinces, demonstrating that in the region of Limpopo Province, especially in upper streams, there is enough radiation in considerations for the development of solar energy projects from a small scale to large scale. The accessibility of the radiation in those fields can be demonstrated to last around twelve hours every day as it is one of the hottest areas in the state.

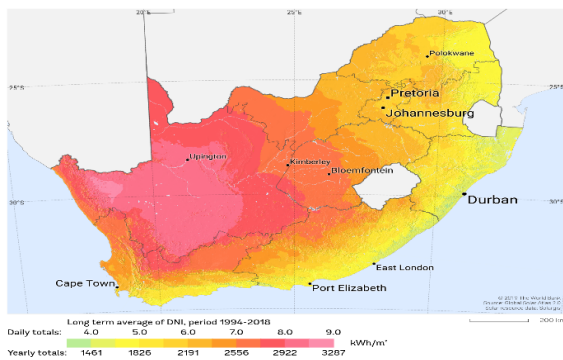


Figure 1. The direct average irradiation through the region (© Global Solar Atlas, 2020)

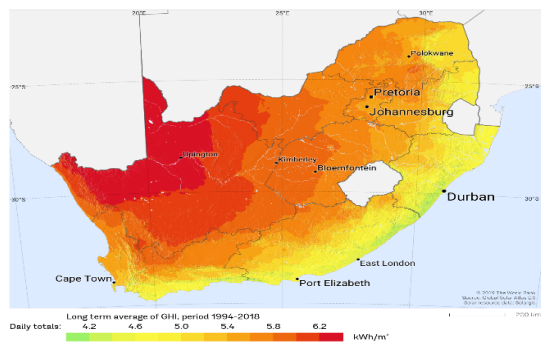


Figure 2. The global horizontal irradiation through the region (© Global Solar Atlas, 2020)

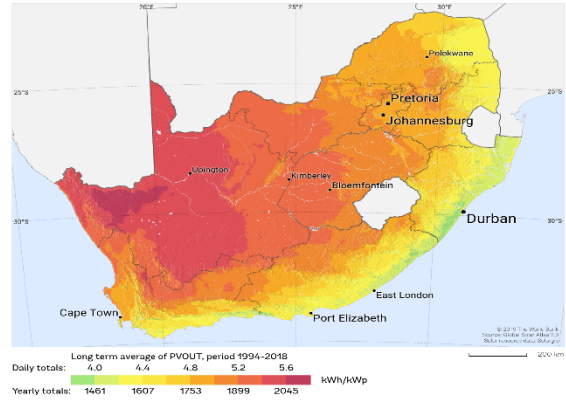


Figure 3. The potential photovoltaic energy for the region (© Global solar Atlas, 2020)

Every country has different solar radiation data but needs other techniques and measurements to determine a specific location. It can be touched on that there are various models produced to determine available solar radiation based on the sunshine hours. In South Africa, mainly, it plays a significant part in renewable energy systems and applications such as in health systems, agriculture, and farming, construction, and housing for domestic and industrial use [1], [28] as it is regarded as the most efficient and economical alternative resource and unused abundant sunshine available throughout the year.

It is essential to know where renewable energy sources come from, as almost all sources originate entirely from the sun [5]. Therefore, the sun's rays that get into the atmosphere are subjected to several elements, including array absorption, scattering, reflection, and transmission through the atmosphere, before arriving at the earth's surface or location of interest. This array is separated into three categories: - diffused, reflected, and direct solar radiation representing the solar energy from the sunlight. Solar radiation data that arrives at the ground level is essential for many wide ranges of applications for meteorology, applied science, environmental sciences, agricultural hydrology and sciences, soil sciences, and physics. In summation, these include the modeling and estimation of crops and crop evapotranspiration, particular health sectors and medicine, and other research in the natural sciences [6], [7], [24].

Many projects are currently being developed, focusing on South Africa and looking into energy systems and their habits. The importance of using solar energy schemes in most parts of Limpopo Province has increased. Independently, improving and raising awareness on issues towards climate change and factors regarding the residential area is one of the significant economic challenges [8]–[10], [27]. This has been understood as the potential importance and opportunity in addressing energy security and carrying out of one the research niche area within the environmental Millennium Development Goals (MDG) on energy [11], [26].

The solar radiation data measured provide information on how much solar energy is reflected along the airfoil during the specific period (i.e., hourly, daily, monthly, and yearly) at the particular position. These measured data are of importance for efficient solar

energy research for utilization and providing energy through natural resources such as photosynthesis [3]-[5]. In modern energy engineering, many cases are calling for equipment subjected to solar radiation [6], [7], [11]. This includes technology equipment such as street lights, traffic lights, remote control gates, public messages and notices, and other solar energy-supplied units installed in the residential district without proper grid-connected electricity supplies. The sunlight power source furnishes this equipment through radiation at defined energy absorption, which causes different significance. Using the radiation data within the Vhembe District allows determining the solar radiation parameters connected to the power needed for solar energy use.

Solar energy forms part of the ultraviolet intensity spectrum, determined by the physical wave solar constant (K). This electromagnetic constant flows through a unit area (A) by the solar array directly to the earth's surface distance from the sun. As a result, to measure the necessary solar energy, it is essential to know the duration time (T) of the solar array path through the aura to the specified destination location [12]. In addition, the amount of incoming solar radiation on the Earth's surface is the measurable amount during the minimum (Tmin). The maximum (Tmax) temperature (that is, in Degree Celsius) at relatively on a daily average of sunshine per hour (hrs); this requires a validated model to be used [12], [26]. Hence the data predictions at the specified location are of importance for the estimation and design in energy conversion for domestic employment industrial and commercial applications [13], [27].

This study introduces a mathematical and estimation radiation model and calculations developed for the Vhembe District to design solar energy strategies as per the radiation data found through the installed meteorological stations. The solar energy is specified and analysed in different geographical locations that enable the parameters in calculations for any positioning on an hourly basis or day by day. As a result, a specific amount of radiation will be employed to determine the possibilities with radiation measurement in planning solar energy systems.

Thus, the mathematical solar energy model used for data prediction provides a vast solar energy potential in the Vhembe District in hourly, daily, and monthly solar radiation on the location determined. Once more, the Vhembe District has a very complex topology, hill mountainous zone area due to its position, and extensive territory with plantations and vegetation, as shown in Figure 4 to Figure 8. As a result, climate data are critical throughout the year due to the location's land cover and weather patterns.



Figure 4. The Thathe Forestry and plantation



Figure 5. The Tshakuma Mango plantation



Figure 6. The Levubu banana plantation



Figure 7. The Elim litchis plantation



Figure 8. The Phiphidi Falls and river

These plantations (including corn, wheat, and sugar cane) furnish the wood wastes through organic materials that can be

applied directly or converted into biofuels or bugs to be burned as fuels to generate energy. Besides, with the availability of the upstream and downstream rivers, there are opportunities that micro/small hydropower systems can be developed to assist the small farmers within the area for the irrigation and plantations, as shown in Figure 8.

In all four districts, people live under a very disadvantageous eco-system and environment, such as limited access to grid-connected electricity and using most natural resources such as traditional biomass, encroachments, paraffin, etc., and gas their energy resources. Also, the district has 14 primary commodities, as shown in Table 1. The majority of the community depends on farming as their economic sustainability, improving their livelihoods, and creation of employment for the rural communities [3], [14], [24].

Table 1. Type of agricultural farmers in the district [2]

Commodities	No. of Smallholder farmers
a) Backyard gardens	644
b) Banana	409
c) Citrus	16
d) Fish	81
e) Garlic	39
f) Guava	128
g) Litchi	4
h) Livestock's	15 652
i) Macadamia	512
j) Mango	758
k) Poultry	992
l) Tomato	2015
m) Vegetable gardens	2300
TOTAL	23 636

With such commodities, an estimation of waste to bio-energy plants could be manufactured to enable the communities to use their waste materials and turn them into a valuable product that can gain them. This will be utilized in determining their energy needs or a marketable product as the source of income in supplying waste to the bio-energy plant that could be got within their area. This will enable the communities to have a whiter, healthier environment and potential job creation and admission to improved energy that improves their living standards. The same uses with biomass, wind and hydro energy resources fail to be taxed.

The measurements received from the solar radiation and humidity climate weather data through the instruments utilized are much subject to stability error function as exposed to heat transport within the aura. This heat transfer is of the drift by 10% of the determined values. In summation, there is a relatively 1% humidity loss of the instruments per month. According to [24], many available studies refer to the global solar radiation models. This includes available models in estimating the daily, weekly, monthly, and annual radiation used for solar energy estimation purposes.

2. Materials and Methods

Throughout the study, the following materials, methods, and analyses were carried out during the estimations and modeling, namely: - Weather data measured throughout the installed metrology weather stations, the GIS maps obtained through Remote Sensing and Mapping downloads for the Vhembe District and using the Photovoltaic graphical modeling through the Matlab software.

Basic Solar Irradiance Measurement

Solar radiation depends on the sunshine that arrives on the earth during the daytime, with its specific latitude location and the atmospheric transmittance (K). Besides the net solar radiation reaching the earth's surface, some are lost and be used for other heating methods, which are turned into additional energy that can be measured using specific instruments [16], [19].

The radiometer is an instrument for measuring irradiance in equal quantities of solar energy at a specified wavelength range measured. The most significant concern was thought in choosing a site placement regarding the determined climatological measurements within the country of interest during the day or in the year. Besides, it was understandable that new models and techniques exist and are being developed in improving the measurement techniques for estimating solar radiation energy with accurate, readable available meteorological parameters. Hence, considering solar radiation on horizontal and tilted surfaces forms part of the estimations [20]. Furthermore, in computing the global radiation, one should take the daily solar radiation absorption (Rs) on the ground, together with the extraterrestrial insolation (Q) and the mean daily solar through the sky transmittance (K) according to equation 1:

$$R_s = Q \times K \tag{1}$$

The constant, K, has a variation of K_c or K_o when there is a clear forecast and K_i on the intermediate days during the year [18], [21]. In summation, in estimating the direct solar radiation (I), one must recognize that it depends on the actual length (r) between the ground and sunlight during the incident measurement. As such, direct solar radiation (I) is known as the dower of the so-called extraterrestrial solar radiation (I_o), which arrives at the earth's surface directly from the atmosphere [20], [24].

Basic Solar Radiation Intensity

The parameters affecting the solar radiation intensity within the atmospheric region are important in solar energy as other arrays are reflected throughout the air. That is, the spectrum of the radiation emitted by the sun is about the power in the ultra-violet region as the solar radiation beam (i.e., constant (I_o)) passes through the atmosphere when the sun is at its mean distance from the earth [21], [22], [25]. This value is

$$I_o = 1.37 \pm 0.02kW/m^2 \tag{2}$$

This constant varies as the light travels through the clouds, absorbed or scattered, reflection and based on the climate latitude and longitude of the location area for the solar energy. This value diverges by 3% as the earth's orbit is elliptical, and the distance from the sun varies all year round. The variations distance between the sun and the world is due to the earth's orbit caused by the actual intensity of solar radiation outside the atmosphere to differentiate

from I_0 by a few per cents to strike into account these variations by a mean factor, F in Degree Celsius.

$$F = 1 - 0.0335 \sin 360(n_d - 94)/365 \quad (3)$$

Where n_d is a specific day of the year (i.e., $n_d = 1$ for January and $n_d = 365$ for December), the argument of the sine function is in degrees. All the values of solar radiation intensity given below, which are in the sun at its average distance from the earth, must be multiplied by F to obtain the actual values on a day. During January, as the weather is clear, the sun is closer to the world, the solar radiation is 3% larger than the average, and in July, when the earth is furthest from the sunshine, the solar radiation is 3% less than the norm.

Prediction Model and Determination of Solar Energy Radiation at Specific Area and Particular Time

The site selection directly impacts the potential renewable energy systems (RES) projects in many different ways, including technical, economic, and environmental aspects. However, one of the critical roles in the PV power plants is the inconsistency and variability of solar irradiation, which can be geographically dissimilar from one location to another [23], [28]. To measure the specific amount of solar radiation at a particular area and a specific time, it is essential to define the angle of inclination by $\cos(\delta)$ of solar arrays to the perpendicular of the earth by considering the area of interest, $\cos(\theta)$ and expressed by the total amount of watts per meters square (W/m^2) and the joules per meter squares (J/m^2) [12], [25]. Also, the RE subject field requires three parameters, namely, global horizontal irradiance (GHI), direct normal irradiance (DNI), and diffuse horizontal irradiance (DHI) [13], [24], [29]. Furthermore, many methods have been introduced to measure global solar irradiation in that respect. These methods have been modified in several ways to suit different models. The simplest example to calculate the global solar radiation is shown in equation 4 [3], [25], 30

$$H = H_o \left[A + B \left(\frac{n}{L_d} \right) \right] \quad (4)$$

The H and H_o are the daily solar radiation and the daily extraterrestrial radiation in $MJm^{-2}d^{-1}$; A and B are constant-coefficient; n and L_d are the sunshine hours per day and location day length in Hours (hrs). The constant-coefficient values are subject to the location of the study and its weather conditions throughout the year [25], [28]. Using the captured weather data, several states are shown using the equation to estimate solar and weather conditions.

In measuring the solar radiation, territorial solar irradiation (E_{ss}) and global solar irradiation (EET) are considered. Hence, the total solar energy (E_s) above the atmospheric level is equal to the absolute atmospheric solar power at sea level multiplied by the length of day (N) per change of temperature (T) throughout the year as determined by equation 5. The latitude (L) and angle of declination (δ) by the sunlight must also be consider parenting the surface of absorption [31].

$$E_s = (I_o + 1) + 0.34 \cos \frac{2\pi N}{365} \times L \quad (5)$$

Where the I_o is the solar constant = $1,367W/m^2$ (is the extraterrestrial radiation as the earth orbits around the sun) and N

is the number of the day for solar absorption, 0.34 being the constant coefficient of solar irradiance at the atmospheric level and L is the length of the day being calculated by equation. The $\cos \frac{2\pi N}{365}$ It is the calculated angle of declination of the sun during the day through per year during the earth's orbit.

The length of the day was calculated by: -

$$L = \frac{2}{15} \cos^{-1}(-\tan L \times \tan \delta) \quad (6)$$

This was determined by the length (L) of days per the solar decline angle (δ) as calculated by equation 7.

$$\delta = 23.45 \sin \frac{(284+N)}{365} \quad (7)$$

In normal circumstances for the Vhembe Region weather measurements, the minimum length was 11.2hrs, and the maximum size of the day was 13.9hrs [3], [25]. Therefore, the solar energy (E_s) for the minimum and maximum was determined by equations 8 and 9.

$$E_{s(\min)} = (I_o + 1) + 3,808 \cos \frac{2N}{365} \quad (8)$$

$$E_{s(\max)} = (I_o + 1) + 4,726 \cos \frac{2N}{365} \quad (9)$$

As a result, equations 1 to 7 were acknowledged in the patterns, estimations, measurements, and computations of the solar radiation energy and demonstrated by the Matlab graphical responses in Figure 17 to Figure 19 for the radiance measurements.

3. Analysis and Discussions

3.1. Meteorology Data Acquisition Analysis

Measurements were remotely captured throughout one year, from January to December 2018. The data acquisition (DAQ) system was used to obtain data from the nine Weather Stations (WS) used during the data collection (i.e., Hanglip, Shefeera, Tsianda, Thohoyandou WO, Dzanani Biaba Agric, Mphefu, Joubertstroom Plantation, Vondo - Bos and Tshivhasie Tea Venda). Table 2 gives the locational longitude and latitude coordinates of the installed weather stations used during the data collection for the study.

Table 2. The location of the weather stations installed in the Vhembe District

Weather Station Name	Longitude (°, E)	Latitude (°, N)
Hanglip	101.07	41.95
Shefeera	94.68	40.15
Tsianda	98.48	39.77
Thohoyandou WO	103.08	38.63
Dzanani Biaba Agric	100.13	37.33
Mphephu	30.03	22.89
Joubertstroom Plantation	22.57	29.19
Vondo - Bos	30.33	23.93
Tshivhasie Tea Venda	22.96	30.35

It was challenging to measure solar radiation in many locations due to the cost of equipment to be used, maintenance, and calibrations to obtain accurate values. Hence, the South African Weather Stations (SAWS) meteorological weather data were used in all the nine weather stations (WS) been installed, in concert with the Agricultural Research Council for the Institute for Soil, Climate and Water (ARC-ISCW) in providing the data. As the results, within reference to the measurements obtained, the data were used to define the solar energy potential for the Vhembe District at specified locations to evaluate its amount for power generation. This data was remotely captured and used to calculate the potential solar energy per location for power generated by photovoltaic system modules. Table 3 shows essential parameters in measuring solar irradiation at different positions. These were applied to evaluate the accessibility of solar irradiation at the designated place.

Table 3: The measurements units used for solar radiation evaluations

Parameters	Units
Global radiation	G (W/m ²)
Diffuse radiation	G _d (W/m ²)
Beam radiation	G _n (W/m ²)
Sunshine hours	σ (hrs)
Maximum and minimum temperature	T _{min} and T _{max} (°C)
Humidity	H (%)
Pressure	P (Pa)
Visibility	F (m)
Wind speed and directions	V(m/s); N,W,E,S
Air mass	ρ(kg/s)

Besides, the efficiency and error calculations were considered difficult to measure the solar radiation at the geographical location due to other factors, including absorbed or reflected by the atmosphere. Following the data obtained, available renewable energy resources within the Vhembe District were of importance and the peoples' quality of life as purpose in using the solar energy to meet their energy demands.

Figure 9 and Figure 10 show the monthly graphical meteorological analysis of the area for minimum and maximum temperature and the length of the day (sunshine hours) as per the yield throughout the year on the direct solar radiation measurement. These measurements were acquired through the day's duration in the district to get a micro solar system for the community expectation and energy demands.

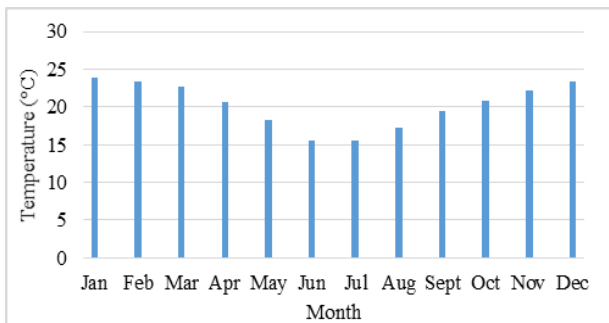


Figure 9: The monthly temperature measurement (Average - 20 °C)

It was noticed that the highest values of the solar insolation are during the summer months (Jan to Apr and Sept to Dec), and the lowest values are during the winter months (May, Jun, Jul, and Aug) as applicable per the day during that time for the season.

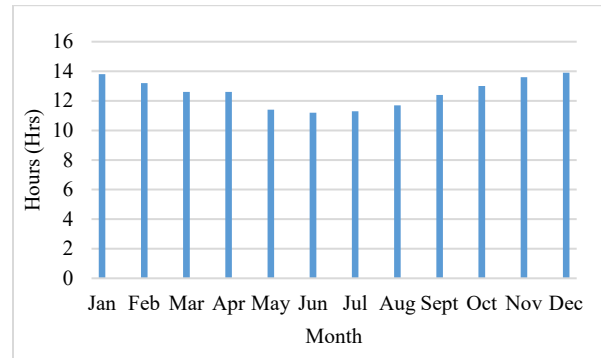


Figure 10: The monthly length of the day measured (Average - 12.5Hrs)

Furthermore, the desirable amount of solar energy was measured during the allowable day time duration (from 06:00 – 19:00) for the whole year per location during the 24hr time interval. The measurements were used to determine the minimum and maximum coefficients variations through the solar radiation calculations using the Matlab software analysis.

3.2. Remote Sensing and GIS Mapping Analysis

The Vhembe District is situated at 22.7696° S Latitude, 29.9741° E 25 Longitude of the Limpopo Province. At an altitude of 250m above mean sea level, a study was conducted in estimating the monthly and annual solar radiation, using the climate and geographical parameters. These areas and the outcomes obtained will assist the researchers and public entities interested in working on solar energy developments to have reference and locations' conditions that they can use for solar energy estimation. Figure 11 shows the potential solar available within the district and per municipality used for domestic use during the solar energy estimation. This map provides an overview of available heat energy to be used.

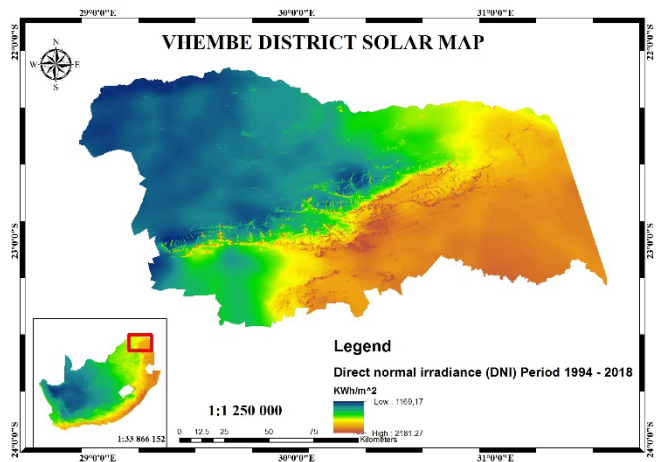


Figure 11: The Solar Map of the Vhembe Strict Area

The analysis shown includes the daily solar radiation assessment from all the locations and using other parameters to see

the solar potential available, as shown by Figure 10 to Figure 13. Thither are many solar radiation databases available for most sections of the countries around the globe.

Collins Chabane, Thulamela, and Makhado Municipality). As a result, the Remote Sensing (RS) for GIS was employed to settle the territorial dominion's solar maps.

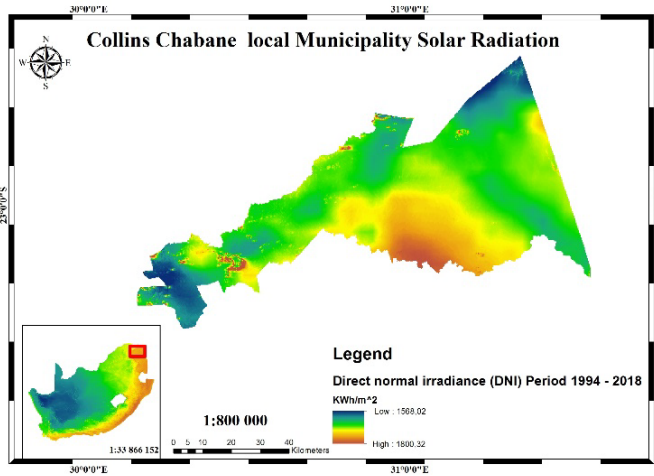


Figure 12: The Areal Solar for the Collins Chabane Municipality

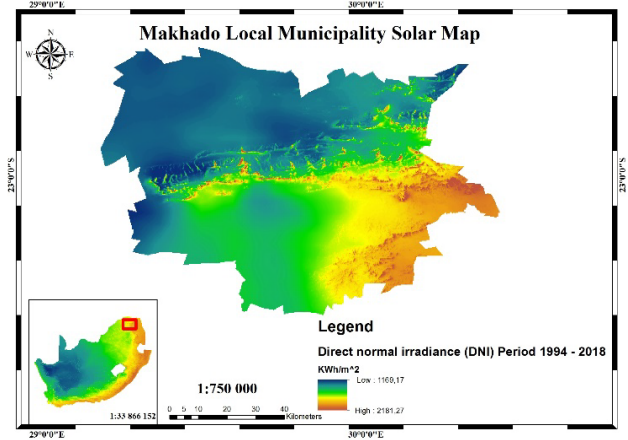


Figure 15: The Makhado Municipality Solar Map

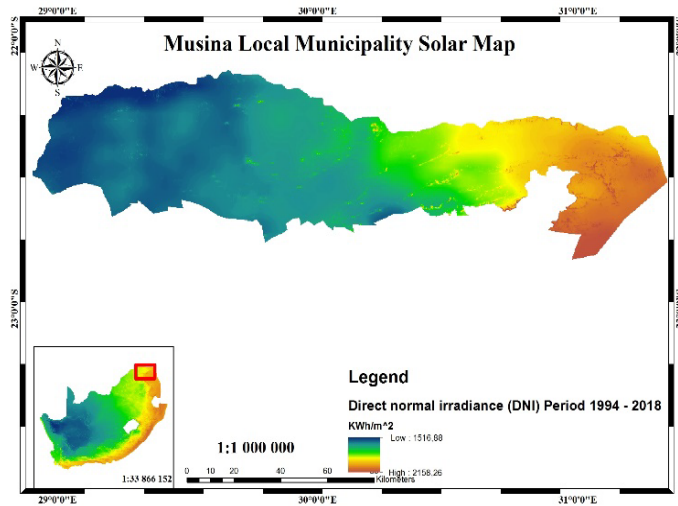


Figure 13: The Musina Municipality Solar Map

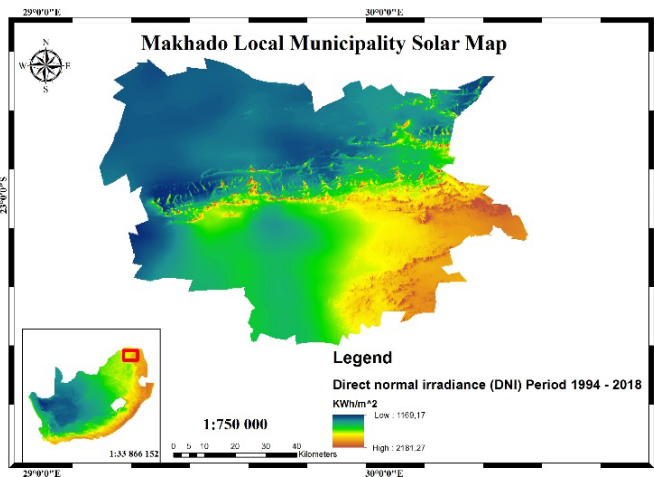


Figure 14: The Thulamela Municipality Solar Map

These were set for the annual weather changes to determine the solar maps of the Vhembe strict and its municipalities (Musina,

3.3. Computational Solar Analysis using Matlab

During the data analysis, the daily solar radiation, the intensity of direct radiation (W/m^2) through an average sun hour of the solar insolation, as shown in Figures 17 to 19. It was noted that high radiation is received during the summertime. There are low irradiance and temperatures in wintertime, which is demonstrated by the lowest and highest measurements for power potential and public presentation.

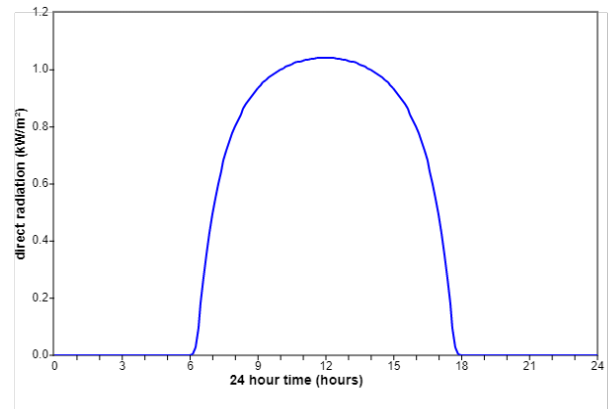


Figure 16: The daily solar radiation for the district

Figure 16 shows the daily solar irradiance curve during the number of hours during sun hours of the day. This is the direct radiation per hour that is generated.

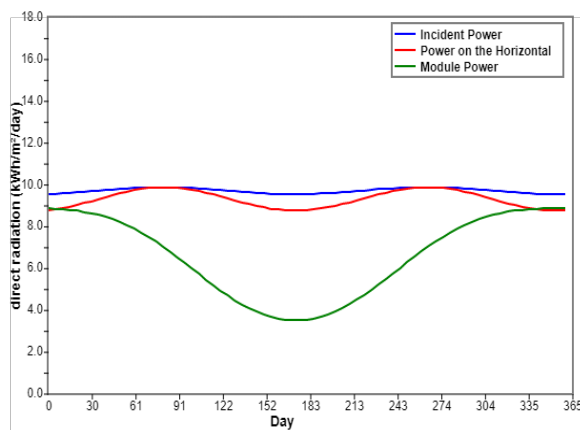


Figure 17: The daily incident power generation per direct intensity

Figure 16 to Figure 18 shows the maximum amount of power directly received without any clouds during the regular sun hours. This amount is set at different to determine how much the specific location's radiation is per period defined. As a result, the power required is generated.

The area is known for its abundant radiation and available solar resources, which significantly influence the design, configuration, and cost of power systems produced. It was observed that the highest values of the solar insolation are during the summer months (Jan to Apr and Sept to Dec), and the lowest values are during the winter months (May, Jun, Jul, and Aug), as demonstrated by Figure 14 to Figure 19. These estimates and weather patterns have been obtained and analyzed through weather stations installed in the Vhembe District and the Matlab software analysis as part of the computations estimations.

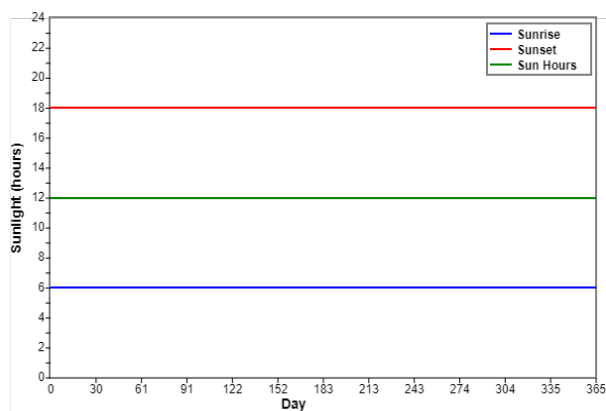


Figure 18: The average sun hours of the solar insolation

Figure 18 shows the mean daily solar availability based on the three curves corresponding to the incoming solar insolation. The daily insolation shown is the number of sun hours rising and sunset. The limited results are helpful for the conception and estimation of the power module needed to take out the solar energy schemes to be set up. This will provide the estimates of the available irradiation generation concept.

The approximation of solar radiation energy is vital in designing the solar energy system or devices. The estimation and size are not easily determined due to the cost and techniques required to practice. As a result, there is a demand for building theoretical methods for estimating solar radiation, such as the

empirical relationships using commonly measured climatological data measured at the specific area [17], [18]. Due to its position and extensive territory, the Vhembe District has a complex topology, hill, and mountainous zone area; hence, the nine installed weather stations found different climate data.

4. Conclusions

The estimation model uses the most recent data measured throughout (January to December 2018) of the meteorological data obtained in the nine installed weather stations. This analysis demonstrates that the required solar radiation values for the potential use in the field are accepted by the temperatures commonly measured by the installed weather stations around the Vhembe District. The obtained results can be the best exemplar for solar estimation at different geographical and climatic locations.

In estimating PV radiation's optimal and potential role, it is essential to consider the location suitability to maximize the solar energy received and the power generated at the selected position. In this paper, the meteorological estimations, graphics, and formulae were applied to influence the behaviour of the Vhembe District climatic conditions as one of the rural regions of involvement in deploying renewable energy technologies such as solar schemes. Consequently, the analysis demonstrates that the required solar radiation values for the potential use in the field are estimated by the temperatures commonly measured by the installed weather stations around the Vhembe District. Furthermore, intensive research studies within specific locations should be carried out to identify and find out the environmental matters linked with the placement and its natural resources to see potential energy sources available for community use.

It is noted that the active use of solar energy schemes has an environmental impact compared to other authors. These solar energy technologies should be increasingly introduced within the rural areas taking into account the suitability and energy potential of the region.

In summary, solar energy provides many advantages over other alternative energy sources. As presented in the paper, a simple principle of solar heat energy can be utilized in various applications. Nonetheless, it is mentioned that solar energy has its drawbacks or limitations like high initial price, depending on the weather, and challenges in energy storage. As a result, the South African Government is increasingly introducing initiatives with plans in providing subsidized programs to increase an effort in encouraging solar energy use in the rural regions. With the application of the solar assessment, the local community will use these findings to assist in determining the potential locations to deploy and install the solar systems for their local use, agriculture, and community use.

Disclosing a conflict of interest

The authors have no conflict of interest to declare.

Acknowledgment

The acknowledgment is towards the support of the Energy Institute (EI) Members and the Centre for Distributed Power & Electronic Systems (DEECE), Dr. K. Aboalez, Dr. M. Adonis, Dr. A. Raji, and Dr. Ali-Mustafa-Ali Almaktoof on their expertise and

supervision in developing, compiling and writing this publication. I want to thank the Research Directorate (RD) Unit under the Office of Deputy-Vice Chancellor Research Innovation and Technology and Partnership (DVC-RITP) for financial support. They appreciate the South African Weather Stations (SAWS) and the Agricultural Research Council for the Institute for Soil, Climate and Water (ARC-ISCW) of data provided. I would likewise like to recognize the 2019 IEEE Power Africa Conference as this extended paper forms part of a conference paper presented at the Abuja, Nigeria conference.

References

- [1] C. Matasane, M.T.E Kahn, "Solar Radiation Estimations Using the Territorial Climatological Measurements in Vhembe District, Limpopo Province for Solar Energy Potential Estimation and Use," 2019 IEEE Power Africa Conf.: Abuja, Nigeria, 2019, doi:10.1109/PowerAfrica.2019.8928806.
- [2] C. Matasane, C. Dwarika, R. Naidoo, "Modelling the Photovoltaic Pump Output Using Empirical Data from Local Conditions in the Vhembe District," 2014 International Conference on Social Education and Community Conf, doi:10.5281/zenodo.1096759.
- [3] M. S. Gadiwala^{1,2}, A. Usman², M. Akhtar², K. Jamil², "Empirical Models for the Estimation of Global Solar Radiation with Sunshine Hours on Horizontal Surface in Various Cities of Pakistan," Pakistan Journal of Meteorology, **9**(18), 2013.
- [4] A. E. Lawin^{1,*}, M. Niyongendako², C. Manirakiza², "Solar Irradiance and Temperature Variability and Projected Trends Analysis in Burundi," Climate 2019, **7**(6), 83, 2019, doi:10.3390/cli7060083.
- [5] S. Zekai, "Solar Energy Fundamentals and Modeling Techniques," Atmosphere, Environment, Climate Change and Renewable Energy, **22**, 2008.
- [6] T. A. McMahon¹, M. C. Peel¹, L. Lowe², R. Srikanthan³, T. R. McVicar⁴, "Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis," Journal of Hydrol. Earth Syst. Sci., **17**, 1331–1363, 2013, doi:10.5194/hess-17-1331-2013.
- [7] M. Paulescu, E. Paulescu, P. Gravila, V. Badescu, "Weather Modeling and Forecasting of PV Systems Operation, Green Energy and Technology," Springer-Verlag London, 17–42, 2013, doi:10.1016/j.rser.2016.11.222.
- [8] P. Jayakumar, Solar Energy Resource Assessment Handbook: APCTT Asian and Pacific Centre for Transfer of Technology of the United Nations – Economic and Social Commission for Asia and the Pacific (ESCAP), 2009.
- [9] L. Mary¹, A. E. Majule², "Impacts of climate change, variability and adaptation strategies on agriculture in semi-arid areas of Tanzania: The case of Manyoni District in Singida Region, Tanzania," African Journal of Environmental Science and Technology, **3**(8), 206–218, 2009, doi:10.5897/AJEST09.099.
- [10] D. R. Brooks, Monitoring Solar Radiation and Its Transmission through the Atmosphere, Department of Mechanical Engineering and Mechanics, Drexel University, Philadelphia, PA, USA, **2**, 2006.
- [11] A Guide to Energy's Role in Reducing Poverty, Energizing the Millennium Development Goals, UNDP, 2005.
- [12] S. A. Kalogirou, "Solar thermal collectors and applications," Progress in Energy and Combustion Science, **30**, 231–295. 2004.
- [13] O. I. Kordun, The influence of solar radiation on sheet steel structures temperature increment, Achieves of Civil Engineering, **LXI** (1), 2015.
- [14] S. S. Ndwakhulu, An evaluation of the performance of the Department of Agriculture in Limpopo Province in improving the livelihood of smallholder farmers during the period 1994–2004, with special reference to the Vhembe District, MSc Thesis, University of Stellenbosch, 2007.
- [15] K. Bakirci, "Models of solar radiation with hours of bright sunshine: a review," Renewable Sustainable Energy Review, **13**, 2580–2588, 2009, doi:10.1016/j.rser.2009.07.011.
- [16] K. Sukarno¹, Ag. S. Abd. Hamid¹, J. Dayou¹, M. Z. H. Makmud², M. S. Sarjadi², "Measurement of Global Solar Radiation in Kota Kinabalu Malaysia," ARPN Journal of Engineering and Applied Sciences, **10**(15), 2015.
- [17] A. A. El-Sebaei, F. S. Al-Hazmi, A. A. Al-Ghamdi, and S. J. Yaghmour, "Global, direct and diffuse solar radiation on horizontal and tilted surfaces in Jeddah, Saudi Arabia," Applied Energy, **87**(2), 568–576, 2010.
- [18] X. Li u, X. Mei, Y. Li, "Evaluation of temperature-based global solar radiation models in China," Agricultural and Forest Meteorology, **149**(9), 1433–1446, 2009, doi: 10.1016/j.agrformet.2009.03.012.
- [19] S. Becker, "Calculation of direct solar and diffuse radiation in Israel," International Journal of Climatology, **21**, 1561–1576, 2001.
- [20] C. K. Pandey and A. K. Katiyar, "Solar Radiation: Models and Measurements Techniques," Journal of Energy, **2013**, ID 305207, doi:10.1155/2013/305207.
- [21] S. Becker; "Calculation of Direct Solar and Diffuse Radiation in Israel," International Journal of Climatology, **21**, 1561–1576. 2001, doi:10.1002/joc.650.
- [22] Z. Samani, "A General Solar Radiation Estimation Model Using Ground Measured Meteorological Data in Sarawak, Malaysia," Journal of Telecommunication, Electronic, and Computer Engineering, **10**(1), 99–105, 2015.
- [23] Z. Hassan, Optimal Design and Analysis of Grid-Connected Solar Photovoltaic Systems, Ph.D Thesis, Concordia University, 2018.
- [24] J. Singh, A. Kruger, "Is the summer season losing the potential for solar energy applications in South Africa?" Journal of Energy South Africa, **28**(2), 2017, doi:10.17159/2413-3051/2017/v28i2a1673.
- [25] S. Bandy, W.A. Zainal, "A General Solar Radiation Estimation Model Using Ground Measured Meteorological Data in Sarawak, Malaysia," Journal of Telecommunication, Electronic, and Computer Engineering, Universiti Teknikal Malaysia Melaka, **10**(1), 99–105, 2018.
- [26] T. R. Govindasamy, N. Chetty, "Quantifying the global solar radiation received in Pietermaritzburg, KwaZulu-Natal to motivate the consumption of solar technologies," Open Physics, **16**(1), 2018, doi:10.1515/phys-2018-0098.
- [27] E.V Tikyaa, A. Akinbolati, M. Shehu, "Assessment of empirical models for estimating mean monthly global solar radiation in Katsina," FUDMA Journal of Sciences (FJS), **3**(1), 333–344, 2019, doi:10.4314/bajopas.v4i2.5.
- [28] Z. E. Mohamed^{*,#}, R. M. Farouk^{**}, H. H. Saleh^{***}, "The Performance Evaluation of Mathematical Models for Predicting MAD-GSR in Egypt," Global Journal of Pure and Applied Mathematics, **14** (7), 897–918, 2018.
- [29] T. S. Mulazdi, N. E. Maluta, and V. Sankaran, "Evaluation of the global solar irradiance in the Vhembe district of Limpopo Province, South Africa, using different theoretical models," Turkish Journal of Physics, **39**, 264–271, 2015, doi:10.3906/fiz-1505-9.
- [30] Ş. Ozan, T. Kuleli. "Estimation of SR over Turkey using artificial neural network and satellite data." Applied Energy **86**, 7–8 (2009): 1222-1228, doi:10.1016/j.apenergy.2008.06.003.
- [31] A. A. El-Sebaei, F. S. Al-Hazmi, A. A. Al-Ghamdi, S. J. Yaghmour, "Global, direct and diffuse solar radiation on horizontal and tilted surfaces in Jeddah, Saudi Arabia," Applied Energy, **87**(2), 568–576, 2010, doi:10.1016/j.apenergy.2009.06.032.

A Study on Novel Hand Hygiene Evaluation System using pix2pix

Fumiya Kinoshita^{*1}, Kosuke Nagano¹, Gaochao Cui², Miho Yoshii³, Hideaki Touyama¹

¹Information Systems Engineering, Graduate School of Engineering, Toyama Prefectural University, Toyama 939-0398 Japan

²Department of Electrical and Computer Engineering, Faculty of Engineering, Toyama Prefectural University, Toyama 939-0398 Japan

³Graduate School of Medicine and Pharmaceutical Sciences, University of Toyama, 930-0194 Japan

ARTICLE INFO

Article history:

Received: 21 January, 2022

Accepted: 09 March, 2022

Online: 18 March, 2022

Keywords:

Hand hygiene

Direct observation method

Generative adversarial networks (GAN)

pix2pix

Discriminant analysis

Mahalanobis distance

ABSTRACT

The novel coronavirus infection (COVID-19), which appeared at the end of 2019 has developed into a global pandemic with numerous deaths, and has also become a serious social concern. The most important and basic measure for preventing infection is hand hygiene. In this study, by photographing palm images of nursing students after hand-washing, using fluorescent lotion to conduct hand-washing training and as a black light, we developed a hand hygiene evaluation system using pix2pix, which is a type of the generative adversarial network (GAN). In pix2pix, the input image adopted was a black light image obtained after hand-washing, and the ground truth image was a binarized image obtained by extracting the residue left on the input image by a trained staff member. We adopted 443 paired-images after hand-washing as training models, and employed 20 images as verification images, which included 10 input images with 65% or more of the residue left, and 10 input images with 35% or less of the residue left in the ground truth images. To evaluate the training models, we calculated the percentage of residue left in the estimated images generated from the verification images, and conducted two-class discriminant analysis using the Mahalanobis distance. Consequently, misjudgment only occurred in one image for each image group, and the proposed system with pix2pix exhibited high discrimination accuracy.

1. Introduction

The novel coronavirus infection (COVID-19), which appeared at the end of 2019, has developed into a global pandemic with numerous deaths, and has also become a serious social concern. COVID-19 is transmitted in the community and develops into a nosocomial infection when an infected person visits the hospital. Because many elderly people and vulnerable patients are hospitalized in hospitals, the damage caused by nosocomial infections is enormous [1]. Developing a novel drug for COVID-19 will take a protracted period of time regardless of the promise of a possible therapeutic drug or the anticipated development of vaccines. Therefore, daily measures against this infection are crucial.

The most important and basic approach to preventing infection is hand hygiene. Hand hygiene is used as a general term that

applies to any form of hand-washing and hand disinfection during surgery, and it refers to washing off organic matter, such as dirt and transient bacteria, that cling to our hands [2]. Although hand-washing has been a cultural and religious practice for centuries, it is believed that the scientific evidence for hand hygiene in the prevention of human diseases emerged only in the early 19th century [3, 4]. The principle of hand hygiene is centered on reducing the dirt and bacteria accumulated on the hands of healthcare workers as much as possible, as well as keeping the hands of healthcare workers clean when they provide care to patients. SARS-CoV-2, the causative virus of COVID-19, invades the body through the mucous membranes of the eyes, nose, and mouth; however, the primary transmission routes that carry the virus to these areas are the fingers [5]. Therefore, keeping the hands clean ensures safety in medical care and nursing, including the safety of both the patient and medical staff. Conventionally, for hand hygiene methods, researchers refer to the guidelines issued by the Centers for Disease Control and Prevention (CDC) and

^{*}Fumiya Kinoshita, Email: f.kinoshita@pu-toyama.ac.jp

World Health Organization (WHO) [6, 7]. These guidelines elucidate hand hygiene methods based on a large number of medical literature and research data, and are easily applicable in actual field situations as they also describe the recommended levels of application [8].

General hand hygiene evaluation methods include the indirect observation method for observing the amount of hand sanitizer, as well as the direct observation method employed for observing the timing for the hand-washing and disinfection method via direct visual inspection. The WHO guidelines recommend the direct observation method by trained staffs [7]. However, this method is limited because it requires trained staff, as well as a significant amount of time to observe several scenes. Additionally, because the direct observation method involves subjective evaluations by the staff, results may vary depending on the skill level of the staff [9]. The methods for the quantitative evaluation of hand hygiene include the palm stamp method, adenosine tri phosphate (ATP) wiping test, and glove juice method [10–13]. The palm stamp method is a method that adopts a special medium, in which the number of bacteria on a hand is visualized by pressing the palm against the special medium, before and after hand-washing. However, this method requires culturing the bacteria present on this hand, which is expensive and time-consuming. Next, the ATP wiping test is a method that analyzes the content of ATP in the living cells of animals, plants, microorganisms, etc. In this method, ATP wiped from the palm is chemically reacted with a special reagent to trigger light emission by the ATP. The amount of light emitted at this point (relative light units; RLU) is evaluated as an index of contamination. Because the ATP wiping test does not require a medium, it is not as expensive and time-consuming as the palm stamp method. However, it is limited because its values fluctuate according to the wiping approach employed with the wiping stick and the amount of strength applied to the wiping stick [14]. Finally, the glove juice method is an evaluation method for hand hygiene recommended by the FDA (U.S. Food and Drug Administration). In this method, the test subject puts on rubber gloves, and the sampling liquid and neutralizer are poured into the gloves. Then, the bacterial liquid in the rubber gloves is cultured, and the amount of bacteria on the hands of the test subject is measured to evaluate hand hygiene. However, the glove juice method is unsuitable for evaluations that focus on each part of the palm as this method evaluates the amount of bacteria present on the entire palm. Therefore, in this study, we developed a hand hygiene evaluation system for evaluating palm images after hand-washing based on the perspective of trained staffs using pix2pix, which is a type of generative adversarial network (GAN).

2. Experimental Method

In the Department of Nursing, Faculty of Medicine, University of Toyama, education on hand hygiene is conducted for nursing students using a hand-washing evaluation kit (Spectro-pro plus kit, Moraine Corporation) [15]. In this process, students apply a special fluorescent lotion (Spectro-pro plus special lotion, Moraine Corporation) on their entire palms and wash their hands hygienically according to the guidelines recommended by WHO. Then, the students themselves evaluate their hand-washing skills by sketching the residue left under a black light. During this hand-washing training, the palm images adopted in this experiment were photographed and used for analysis. This experiment was

conducted after obtaining approval from the Toyama Prefectural University Ethics Committee.

A special imaging box (dimensions: 30 × 30 × 45 cm) was developed to merge the shooting environment for the palm images. Two LED fluorescent lights (RE-BLIS04-60F, Reudo Corp.) and two black lights (PL10BLB, Sankyo Denki) were installed inside the imaging box (Fig. 1). Although the black light was turned on while shooting the palm image, the entrance part was shaded with a blackout curtain, and then the picture was photographed using a digital camera (PowerShot G7 X Mark II, Canon). The number of pixels of the captured image was 5472 × 3072 px, and the parameters at the point of shooting, including resolution, F value, shutter speed, ISO sensitivity, and focal length were 72 ppi, 2.8, 1/60 s, 125, and 20 cm, respectively. For these parameters, the numerical values that are in focus on the palm without overexposing the image, even when shooting with the fluorescent light on, were selected. The procedure for obtaining the palm image is provided below.

- The test subject applies the fluorescent lotion on the entire palm and confirms that the fluorescent lotion is applied correctly under the black light.
- The test subject performs hygienic hand-washing using running water and soap according to WHO guidelines. After hand-washing, the test subject wipes off moisture with a paper towel.
- The test subject then places the palm of their right arm upwards in the imaging box, and the photographer takes one image each under the fluorescent and black lights (Fig. 2).

Via the hand-washing training, 463 palm images after hand-washing were obtained in this experiment.



Figure 1: Illustration of imaging box

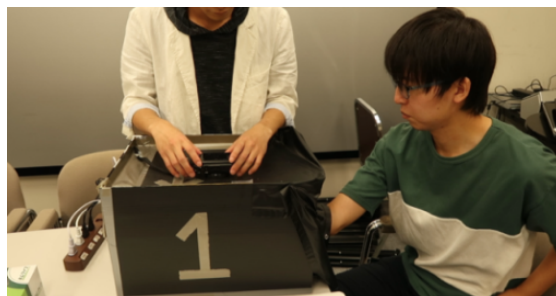


Figure 2: Experimental scene

3. Generation of training models using pix2pix

3.1. Characteristics of pix2pix

In recent years, machine learning technology has been remarkably improved and applied to various fields such as natural language processing and speech recognition, as well as imaging areas such as image classification, object detection, and segmentation [16]. Among these applications, generative adversarial networks (GANs) have been garnering considerable attention recently. GAN is an unsupervised learning method for neural networks proposed by [17]. It is characterized by its adoption of two convolutional neural networks called generator and discriminator in opposition. The generator learns to generate images such that generated images are not misidentified as images generated by the discriminator, and the discriminator learns such that the generated data are not misidentified as the training data. Accordingly, an image similar to the training data is ultimately generated. The pix2pix adopted in this study is also a type of GAN, as well as an image generation algorithm proposed by [18]. The pix2pix learns the relationship between images from two paired images: input and ground truth images, which generate an estimated image considering the relationship obtained when inputting the input image. In previous studies, pix2pix has been used to generate a wide variety of images, such as creating maps from aerial photographs and generating color images from black-and-white images [18–20]. Conventionally, it is necessary to design a network to address a specific challenge when generating images. However, in the proposed method, it is unnecessary to design a network for each problem as pix2pix is a highly versatile image generation algorithm. Implementation methods that can easily manage pix2pix are also published on the internet, and these methods can be simply adopted by preparing two paired images: the input and ground truth images. In this study, we implemented pix2pix according to the pix2pix-Tesorflo [21] repository on GitHub.

3.2. Pre-processing of training data

The input image adopted for pix2pix in this study was a black light image obtained after hand-washing, while the adopted ground truth image was a binarized image obtained by extracting the residue left from the input image by a trained staff from a faculty in the Department of Nursing. The pre-processing procedure for the paired image is presented below.

- Dirt, such as the adhesion of fluorescent lotion to the floor surface of the imaging box, may affect the analysis. Therefore, using the image captured under fluorescent light, only the right arm part was extracted from the image captured under the black light. In contour extraction, MATLAB's image processing toolbox was employed to perform edge detection and basic morphology via Sobel's method [22]. Fig. 3a presents an image captured under the fluorescent light, while Fig. 3b illustrates a binarized image in which the outline of the arm area is extracted from Fig. 3a. In addition, the arm area is painted white and the rest of the area is painted black. Fig. 3c presents the results of superimposing Fig. 3b on the image captured under black light.
- In this system, the parts from the arm to the wrist of the black light image with the background (painted in black) is excluded

to focus on the residue left in the palm area (Fig. 4a). It is necessary to change the length-width pixel ratio of the input image of pix2pix to 1:1. Accordingly, the black pixels were evenly arranged at the top and left-right corners of the image such that the number of pixels of all black light images became 4746×4746 px. After that, the residue left was extracted visually from the black light image by the trained staff. In the extraction of the residue left, the brightness value of the black light image was targeted, and binarization image processing was conducted such that the areas the trained staff felt were left unwashed are displayed in black (Fig. 4b). Furthermore, only the residue left is depicted as black in the actual ground truth image because the training model did not include the contour extraction of the palm.

The above-mentioned processing method was carried out on all 463 captured images. Among these images, 20 images, which include 10 input images with 65% or more of the residue left and 10 input images with 35% or less of the residue left in the ground truth images, were randomly extracted to evaluate the training models, and these images were adopted as verification images. When generating the training models, the paired images were reduced to an image size from 4746×4746 px to 256×256 px. The other parameters considered, which include batch size and learning rate values were 10 and 0.00002, respectively. In addition, in the 443 paired images adopted in the training models, a bias exists in their percentage of residue left in the ground truth images. Therefore, we also investigated the training models when a data expansion technique (data augmentation) is applied, such that the percentage of residue left in the ground truth images is uniform [23].

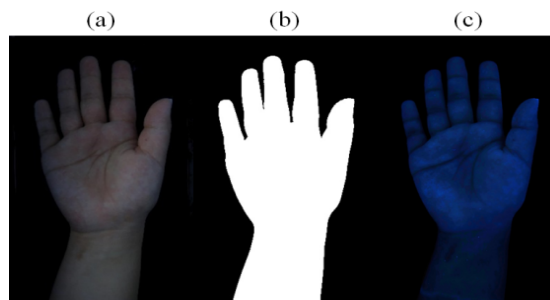


Figure 3: Pre-processing method of input image

- (a) Image captured under fluorescent light, (b) Binarized image, in which arm area is extracted from (a), (c) Image captured under black light, in which every area other than the arm area is painted in black using (b)

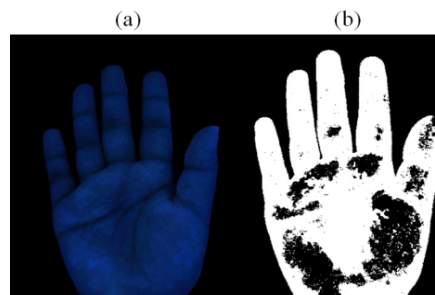


Figure 4: Method for generating ground truth image

- (a) Input image (brightness tripled for the paper), (b) Binarized image, in which the residue left is extracted from the input image

4. Results

In this experiment, 463 palm images were obtained, and among them, 20 were used as verification images, such that the remaining 443 paired images were adopted as pix2pix training models. In addition, a histogram was created from the 443 ground truth images, in which the horizontal and vertical axes represent the percentage of residue left (in increments of 10%) and number of images, respectively. Furthermore, we also examined the training models at the point where a negligible bias appeared in the percentage of residue left in the ground truth images. In data augmentation, which is a data expansion technique, three processes, including enlargement, reduction, and translation, were conducted on the palm images. In both the enlargement and reduction processes, the size of the palm was altered from 85% to 115% in 5% increments. In the translation process, the position of the palm was moved horizontally by 5 px and 10 px, respectively. Fig. 5 presents a histogram of the percentage of residue left adopted in the training models, and the number of images. Fig. 5a presents a histogram of the number of images present in the 443 paired images while Fig. 5b presents a histogram of 1800 paired images after data augmentation.

An estimated image was generated from the verification image using each training model, and the percentage of residue left was calculated accordingly (Fig. 6). Fig. 6a presents the percentage of residue left in the 443 training models while Fig. 6b presents the percentage of residue left in the training models after data augmentation. Subsequently, a two-class discriminant analysis was performed using the Mahalanobis distance as the percentage of residue left in the estimated image [24]. Consequently, the misjudgment in the training models without data augmentation occurred in two images for each image group, and the misjudgment in the training models with data augmentation occurred in one image for each image group. Examples of the input, ground truth, and estimated images adopted for evaluation in each training model are presented in Figs. 7–9 (note that the input image is tripled in brightness for this paper). Fig. 7 presents an example in which the percentage of residue left in the estimated image was correctly discriminated in both training models. Fig. 8 illustrates an example in which the percentage of residue left in the estimated image was solely correctly discriminated in the training model after data augmentation. Finally, Fig. 9 shows an example in which the percentage of residue left in the estimated image could not be correctly discriminated in both training models.

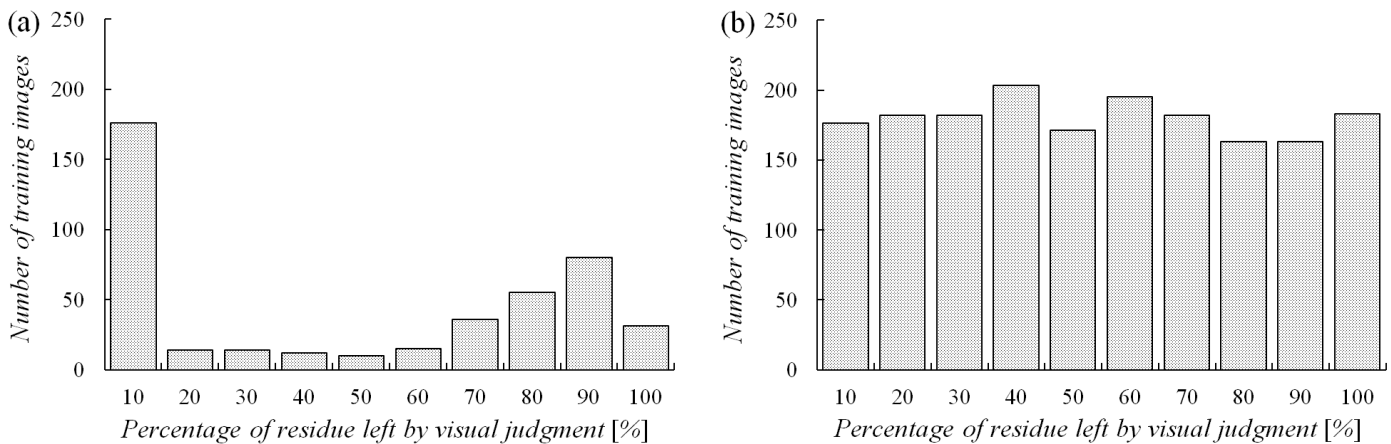


Figure 5: Histogram of percentage of residue left in the ground truth image in each training model

(a) Training model without data augmentation, (b) Training model with data augmentation

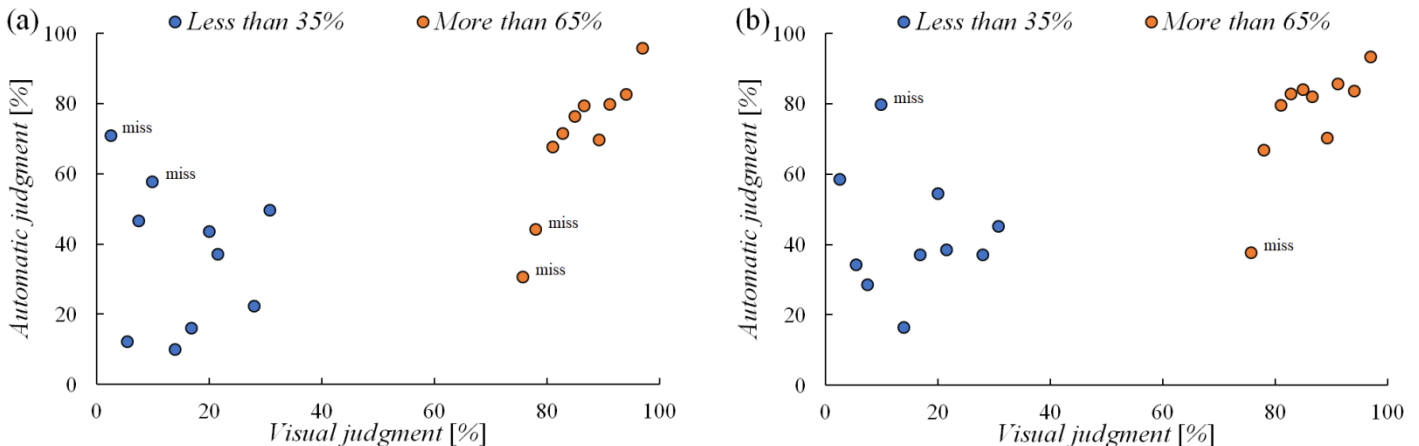


Figure 6: Percentage of residue left in the estimated image generated from the verification image

(a) Training model without data augmentation, (b) Training model with data augmentation

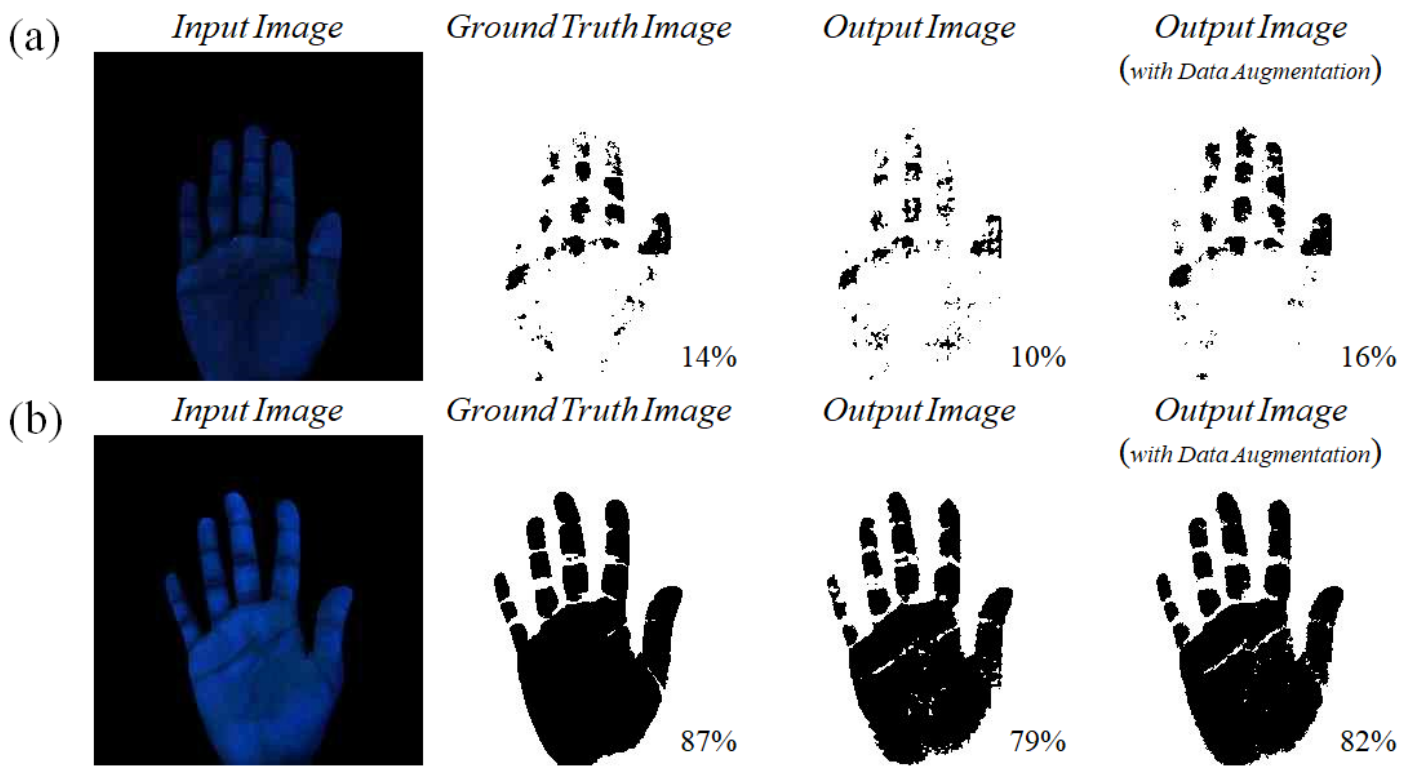


Figure 7: Examples of verification images correctly discriminated in both training models
 (a) Input image with 35% or less of residue left, (b) Input image with 65% or more of residue left

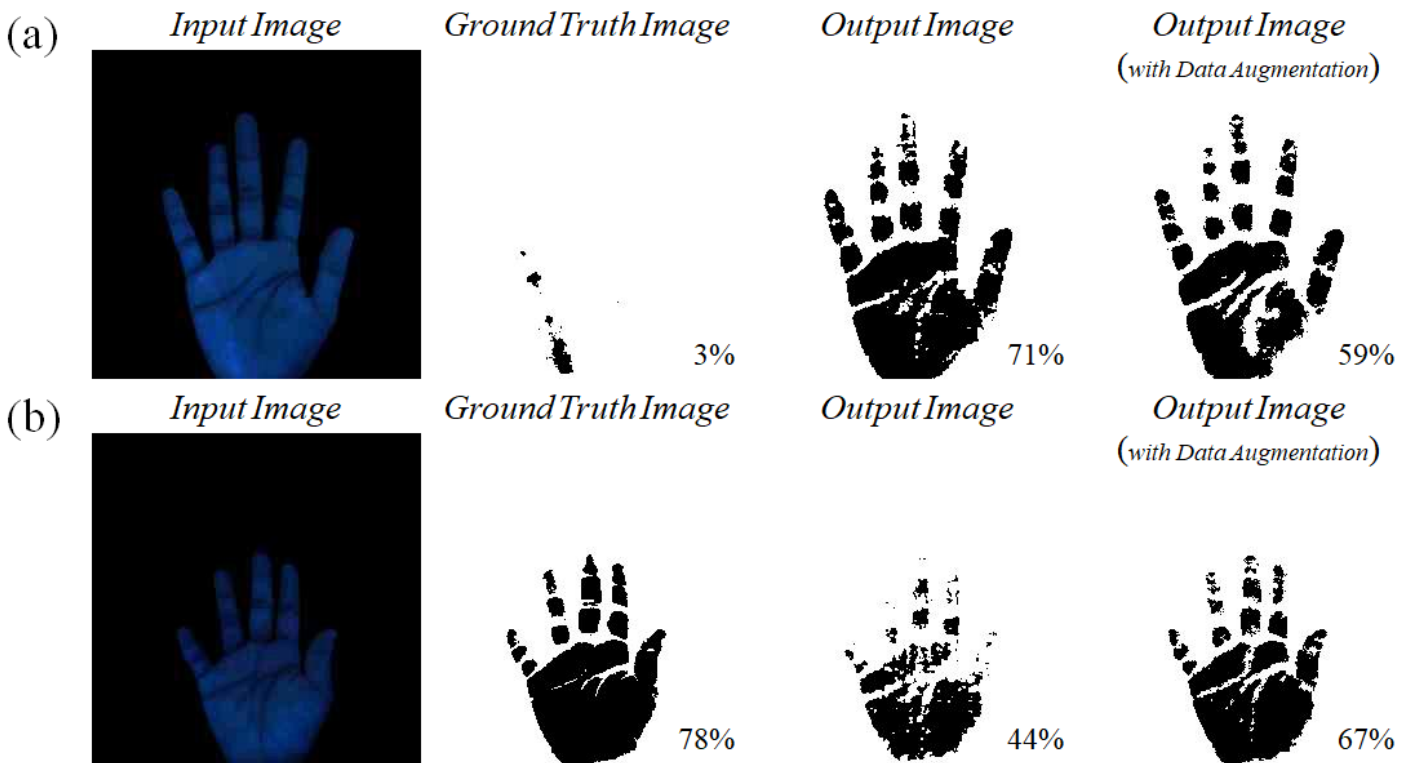


Figure 8: Examples of verification images that were correctly discriminated only in the training model with data augmentation
 (a) Input image with 35% or less of residue left, (b) Input image with 65% or more of residue left

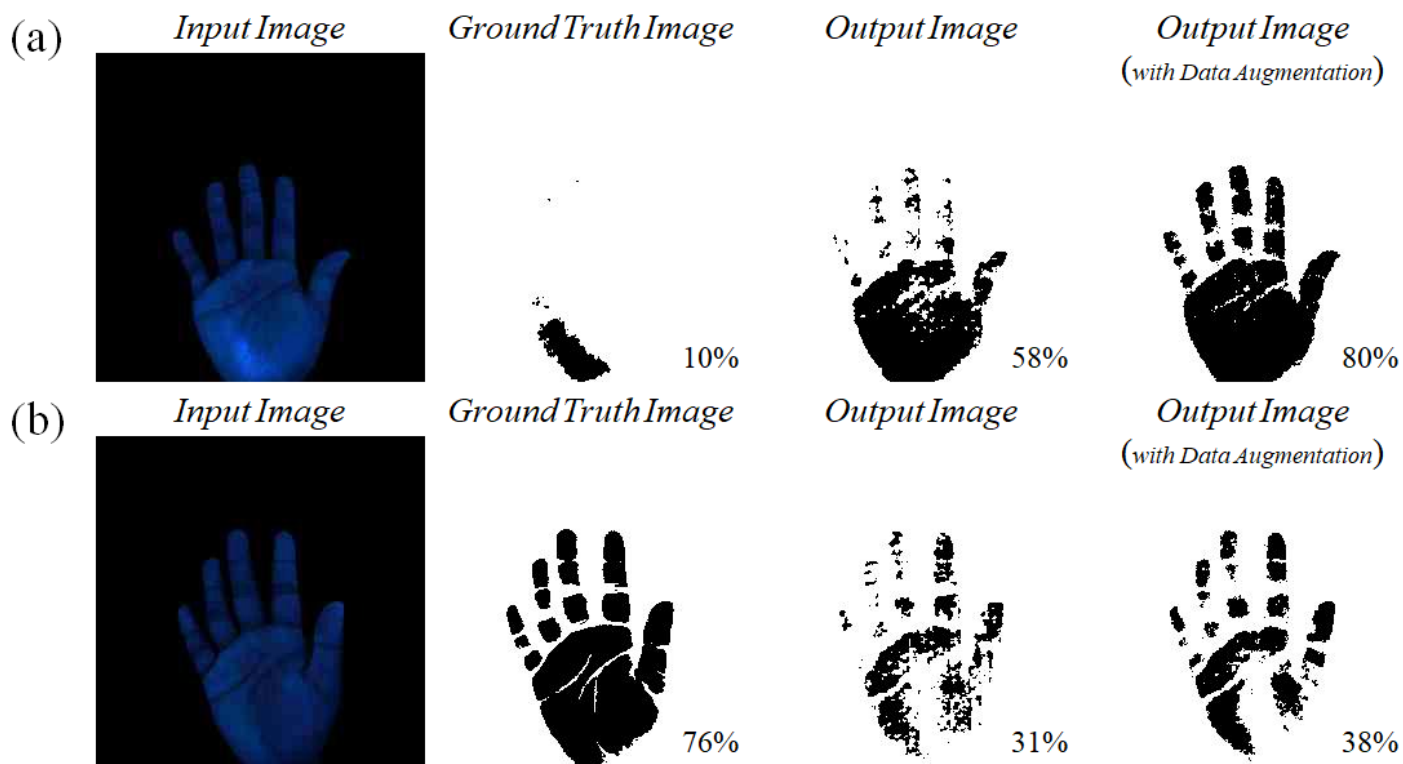


Figure 9: Examples of verification images that could not be correctly discriminated in both training models

(a) Input image with 35% or less of residue left, (b) Input image with 65% or more of residue left

5. Discussions

WHO guidelines recommend direct observation by trained staffs as a general hand hygiene evaluation method. However, certain factors have been identified as limitations to this method, such as the requirement to secure trained staff and the amount of time required to observe several scenes in this method. In addition, because the direct observation method involves subjective evaluations by the staff, results obtained may vary depending on the skill level of the staff. Therefore, based on the perspective of trained staffs, we developed a hand hygiene evaluation system in this study to evaluate palm images after hand-washing using pix2pix, which is a type of generative adversarial network (GAN). The input image adopted for pix2pix was a black light image obtained after hand-washing, and the ground truth image employed was a binarized image obtained by extracting the residue left from the input image by a trained staff. Regarding the generated estimated image, an image similar to the ground truth image was generated when the verification image was input to the pix2pix training model. In addition, the discrimination accuracy of the verification image was improved in the training models via the data augmentation method. The percentage of residue left was 14% in the ground truth image of Fig. 7a, whereas it was 10% in the estimated image. In contrast, the percentage of residue left for the estimated image was improved to 16% via the data augmentation method. Similarly, the percentage of residue left was 87% in the ground truth image of Fig. 7b, whereas it was 79% in the estimated image. The percentage of residue left in the estimated image was improved to 82% via the data augmentation method

Fig. 8 presents an example in which the percentage of residue left in the estimated image was only correctly discriminated in the training model after data augmentation. In both Figs. 8a and b, the percentage of residue left in the estimated image exhibited a tendency to approach the percentage of residue left in the ground truth image via the data augmentation method. In this pix2pix method, it was verified that the learning rate is improved by eliminating the bias in the percentage of residue left in the ground truth image of the training models. In contrast, Fig. 9 presents an example in which the percentage of residue left in the estimated image could not be discriminated correctly in both training models. In Fig. 9b, the percentage of residue left in the estimated image exhibited a tendency to approach the ground truth image via the data augmentation method; however, in Fig. 9a, it exhibited a tendency to move away from the ground truth image. There are two possible reasons for this phenomenon. First, the palm image in Fig. 9a might have been insufficiently shaded at the time of shooting. After inserting the right arm, a blackout curtain was adopted to shade the entrance at the time of shooting. However, the parts without the light emission of the fluorescent lotion were emphasized owing to insufficient shading, which may be ascertained to be the residue left. Second, it is possible that the training model could not be optimally created in the training data with a small percentage of residue left. In the training data with a significant percentage of residue left, the entire palm is painted black, such that the effect of residue left is negligible owing to these parts. However, in the training data with a negligible percentage of residue left, the position of the residue left differs for each part, which indicates the possibility of insufficient training. Fig. 6 presents the percentage of residue left in the verification and estimated images. In the image group with a

significant percentage of residue left, the variation in the data is negligible, and the estimated image often takes a value approximate to that of the ground truth image. However, in the image group with a negligible percentage of residue left, the variation in the data is significant and the values are also scattered. In other words, it is necessary to further increase the number of data used for training data with a small percentage of residue left. However, in the two-class discriminant analysis with the Mahalanobis distance, the misjudgment of both image groups occurred in one image each, and the analysis exhibited high discriminant accuracy.

6. Conclusions

Several infection control measures have been adopted in the medical field to address the unprecedented infectious disease called the novel coronavirus. Among them, the most important and basic approach to preventing infection is hand hygiene. Conventionally, for hand hygiene methods, researchers refer to the guidelines issued by CDC and WHO. However, the direct observation method, which is a subjective method, has been proposed as the gold standard for hand hygiene evaluation. Therefore, in this study, we developed a hand hygiene evaluation system using pix2pix as a simple and quantitative hand hygiene evaluation method. In this system, using the ground truth image of pix2pix from the perspective of trained staff, we addressed challenges such as securing trained staff and the protracted time required for evaluation, which are the limitations to the conventional direct observation method. Furthermore, in the two-class discriminant analysis for the presence or absence of residue left, misjudgment occurred in one image each for both image groups, and the analysis exhibited high discriminant accuracy. In the future, we will aim to apply the proposed system as the primary screening for hygienic hand-washing by further increasing the training data and improving discrimination accuracy. In addition to the medical field, we will apply this system in the educational field by investigating methods that do not require fluorescent lotion or black light.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] T. Kikuchi, "COVID-19 outbreak: An elusive enemy," *Respiratory Investigation*, **58**(4), 225–226, 2020, doi: 10.1016/j.resinv.2020.03.006.
- [2] H. Misao, "Hand hygiene concept: unchanging principles and changing evidence," *The Journal for Infection Control Team*, **11**(1), 7–12 2016.
- [3] J. D. Katz, "Hand washing and hand disinfection," *Anesthesiology Clinics of North America*, 2004, doi: 10.1016/j.atc.2004.04.002.
- [4] A. G. Labarraque, "Instructions and Observations concerning the use of the Chlorides of Soda and Lime," *The American Journal of the Medical Sciences*, 1831, doi: 10.1097/00000441-183108150-00021.
- [5] K. Yano, "Novel coronavirus (COVID-19) infection measures," *Infection and Antimicrobials*, **23**(3), 176-181, 2020.
- [6] J. M. Boyce, D. Pittet, "Guideline for hand hygiene in health-care settings: Recommendations of the Healthcare Infection Control Practices Advisory Committee and the HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force," *American Journal of Infection Control*, 2002, doi: 10.1067/mic.2002.130391.
- [7] World Health Organization, "WHO guidelines on hand hygiene in health care," World Health, 2009.
- [8] M. Ichinohe, "Basics of hand hygiene technique: how to choose a method that suits the on-site situation," *Journal for Infection Control Team*, **11**(1), 19–25, 2016.
- [9] V. Erasmus et al., "Systematic Review of Studies on Compliance with Hand Hygiene Guidelines in Hospital Care," *Infection Control & Hospital*

- Epidemiology*, **31**, 283–294, 2010, doi: 10.1086/650451.
- [10] T. Kato, "Systematic Assessment and Evaluation for Improvement of Hand Hygiene Compliance," *Journal of Environmental Infections*, **30** (4), 274–280, 2015.
- [11] M. Sato, R. Saito, "Nursing Students' Knowledge of Hand Hygiene and Hand Hygiene Compliance Rate during On-site Clinical Training," *Japanese Journal of Environmental Infections*, **34**(3), 182–189, 2019.
- [12] U.S. Food and Drug Administration (FDA), "Guidelines for effectiveness testing of surgical hand scrub (global juice test)," *Federal Register*, **43**, 1242–1243, 1978.
- [13] Y. Hirose, H. Yano, S. Baba, K. Kodama, S. Kimura, "Educational Effect of Practice in Hygienic Handwashing on Nursing Students: Using a device that allows immediate visual confirmation of finger contamination," *Environmental Infections*, **14**(2), 123–126, 1999.
- [14] K. Murakami, S. Umesako, "Study of Methods of Examination by the Full-hand Touch Plate Method in Food Hygiene," *Japanese Journal of Environmental Infections*, **28**(1), 29–32, 2013, doi: 10.4058/jsei.28.29.
- [15] M. Yoshii, K. Yamamoto, H. Miyahara, F. Kinoshita, H. Touyama, Consideration of homology between the visual sketches and photographic data of nursing student's hand contamination, *International Council of Nursing Congress*, 2019.
- [16] T. Shinozaki, "Recent Progress of GAN: Generative Adversarial Network," *Artificial Intelligence*, **33**(2), 181–188, 2018, doi: 10.11517/jjsai.33.2_181
- [17] I. J. Goodfellow et al., "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2672–2680, 2014, doi: 10.3156/jsoft.29.5_177_2.
- [18] P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 5967–5979, 2017, doi: 10.1109/CVPR.2017.632.
- [19] Y. Taigman, A. Polyak, L. Wolf, "Unsupervised cross-domain image generation," 2017.
- [20] M. Y. Liu, T. Breuel, J. Kautz, "Unsupervised image-to-image translation networks," in *Advances in Neural Information Processing Systems*, 701–709, 2017.
- [21] <https://github.com/tensorflow/docs/blob/master/site/en/tutorials/generative/pix2pix.ipynb>
- [22] <https://jp.mathworks.com/help/images/detecting-a-cell-using-image-segmentation.html?lang=en>
- [23] C. Shorten, T. M. Khoshgofaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, **6**(1), 2019, doi: 10.1186/s40537-019-0197-0.
- [24] G. J. McLachlan, "Mahalanobis distance," *Resonance*, **4**(6), 20–26, 1999, doi: 10.1007/bf02834632.

A Unified Visual Saliency Model for Automatic Image Description Generation for General and Medical Images

Sreela Sreekumaran Pillai Remadevi Amma^{*1}, Sumam Mary Idicula²

¹Department of Computer Science, Government College Kariavattom, Thiruvananthapuram, Kerala, India

²Department of Computer Science, Muthoot Institute of Technology and Science, Kochi, Kerala, India

ARTICLE INFO

Article history:

Received: 21 January, 2022

Accepted: 09 March, 2022

Online: 28 March, 2022

Keywords:

Image Description Generation

Image captioning

Deep learning

Visual Attention

ABSTRACT

An enduring vision of Artificial Intelligence is to build robots that can recognize and learn the visual world and who can speak about it in natural language. Automatic image description generation is a demanding problem in Computer Vision and Natural Language Processing. The applications of image description generation systems are in biomedicine, military, commerce, digital libraries, education, and web searching. A description is needed to understand the semantics of the image. The main motive of the work is to generate description of the image using visually salient features. The encoder-decoder architecture with a visual attention mechanism for image description generation is implemented. The system uses a Densely connected convolutional neural network as an encoder and Bidirectional LSTM as a decoder. The visual attention mechanism is also incorporated in this work. The optimization of the caption is also done using a Cooperative game-theoretic search. Finally, an integrated framework for an automatic image description generation system is implemented. The performance of the system is measured using accuracy, loss, BLEU score and ROUGE. The grammatical correctness of the description is checked using a new evaluation measure called GCorrect. The system gives a the-state-of-art performance on the Flickr8k and ImageCLEF2019 challenge dataset.

1. Introduction

Computer vision researchers nowadays mainly focus on descriptive language to describe the world. Image description generation is a challenging topic in computer vision and natural language processing. Deep learning technology has produced tremendous progress in the automatic generation of image descriptions. The image description expresses the semantics and linguistic representation of the image. Image captioning applications include image retrieval, automatic video surveillance, image indexing, education, aid to visually impaired people, etc. The caption of a photo contains objects, attributes, spatial relationships, and actions. The description of an image should be meaningful, self-contained, grammatically, and semantically correct.

Generating image descriptions is an essential process in the area of both Computer Vision and Natural language processing. Imitating the human attitude for giving descriptions for images by a machine is a noticeable step along with the rapid growth of Artificial Intelligence. This task's significant challenge is to capture the relationships of objects in an image and generate a

natural language description. Traditionally, predefined templates are used for generating descriptions. However, this approach does not give enough variety available for creating lexically and semantically detailed descriptions. This limitation has been conquered with the increased efficiency of neural network models. Neural networks generate captions in state-of-the-art models by giving an image as input and forecasting the output description. The automatic image description generation system has many critical applications, such as aiding visually impaired people, building an intelligent robot, and making Google Image Search better than Google Keyword Search.

Connecting image and language is a complex problem in Computer Vision and Natural language Processing. Based on the literature, a comprehensive scene understanding is difficult. The image description systems should produce grammatically correct, relevant, human-like, and describe accurate information. For generating a better caption, the vital image features should be selected. Content selection from images is a significant problem. To optimize the description, the max search and beam search are commonly used methods for determining words. So, caption optimization needs to be improved to eliminate the limitations of

* Corresponding Author Sreela S R, Email: sreela148@gmail.com

max search and beam search.

Analyzing and summing up ideas from clinical pictures such as radiology images is a tedious task that specialists can handle. Automatic methods approximate the mapping from visual information to condensed textual descriptions. All medical images training data are accompanied by UMLS concepts extracted from the original image caption. Medical image captioning is an actual application of automatic image description generation. In this work, the proposed automated image description system is used for medical image captioning.

- We developed two components as a visual attention architecture and integrated automatic image description generation system to achieve the objectives. They are explained below:
- Visual attention: A hybrid architecture for visual attention is implemented. Spatial, channel-wise, and layer-wise attention are the components of visual attention.

Integrated automatic image description generation system: The image features are extracted using Densenet. The description is generated using Bidirectional Long Short term memory (BLSTM). The caption optimization is implemented using game-theoretic search. The framework has experimented on the Flickr8k dataset and medical image dataset ImageCLEF2019.

2. Related Works

Recent trends in Computer Vision and Natural language processing have influenced image description generation. An automatic image description generation is essential for many reasons, such as image understanding, image indexing, image searching, etc. Many research works have been progressed in image description generation in the last ten years. Image captioning systems are classified into Traditional machine learning-based systems and deep learning-based systems. In our literature, we concentrate more on deep learning-based approaches. The taxonomy of deep learning-based image description generation is based on six criteria: type of machine learning approach, the model architecture, feature mapping, the language models used, the number of captions produced, and others. The model architectures used in this system are encoder-decoder architecture and compositional architecture. The features are mapped into two spaces, such as visual space and multimodal space. The language models used in this system are Recurrent Neural Network(RNN) [1], Long Short Term Memory(LSTM) [2], etc. The captioning systems are divided into dense captioning and scene-based captioning based on the number of captions generated. Other image description generation systems are attention-guided, semantic concept-oriented, novel object-based, and stylized captions.

From the literature, the image description generation systems are classified as follows.

Direct Generation model: This model extracts the image's visual content and generates a description. Google's Neural Image caption generator [3], BabyTalk [4], Midge [5], Karpathy's system [6] follow this model.

Visual space model: The visual model finds the identical images of the query image and maps the description to the image.

Multimodal space model: This model finds similar images from multimodal space such as visual and textual. This kind of image description generation system is considered a retrieval problem.

The image captioning systems are further classified into template-based systems and Deep Neural Network-based systems.

Template-based: In this approach, the captions are generated using the objects, attributes, and scenes. Farhadi et al. The Markov random field, GIST, and Support Vector Machine(SVM) are used for caption generation and transform the scene contents into natural language sentences using the template in [7] systems. The Conditional Random Field (CRF) relates the objects, attributes, and prepositions systems [4]. Midge [5] systems generate text using the Berkeley parser and Wordnet ontologies. The disadvantage is that they produce inaccurate descriptions due to wrong object detection. Classical machine learning algorithms are used for object detection, which results in bad performance.

Deep Neural Network-based approach: The image captioning system involves image to text translation. Currently, the image captioning system consists of two parts: Encoder and Decoder. The encoder is used for extracting features of the picture. Deep neural networks are used for encoding, which has the highest accuracy in object categorization. The decoding module is realized using recurrent neural networks or LSTM, which is practiced for caption generation. The main components in [8] system are fully connected neural networks and multimodal log-bilinear models. A few works worked on the recurrent neural network for caption generation. In [9], the system generate image descriptions using deep CNN(Convolutional Neural Network) and bag of words. Karpathy[10] produces a dense description of images using Region level CNN (RCNN) and bidirectional RNN. In [3], it is identified LSTM gives better performance for decoding operation. The system in [11] map the relationship between learned word embeddings and the LSTM hidden states. Authors generate captions using a deep Convolutional Neural Network (CNN) and two distinct LSTM networks for analyzing forward and backward direction of description. The approaches used in [12] systems are top-down and bottom-up.

3. Proposed System

Figure 1 represents the detailed architecture of the proposed system. In the proposed method, the system's primary goal is to enhance the automatic image description generation system's efficiency by generating meaningful sentences. The objects, attributes, actions, scenes, etc., are treated as image features. The sentence features are Noun, Adjective, verb, preposition, etc. The image captioning system maps the image features to sentence features. The essential tasks in this system are image parsing, sentence modelling, and surface realization. Preprocessing, feature extraction, and visual attention are the crucial steps in image parsing. Sentence modelling contains preprocessing and Text encoding. Caption generation, optimization, and Evaluation are the critical steps in surface realization.

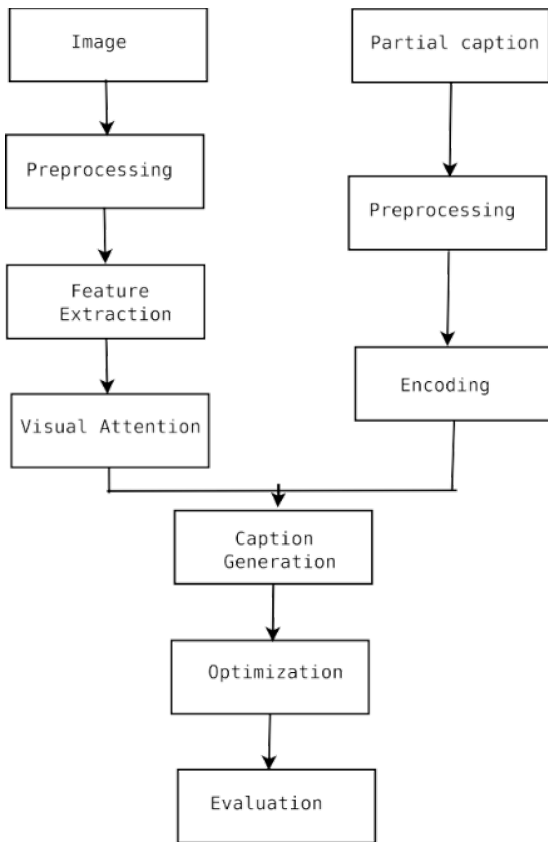


Figure 1: Methodology of Proposed System

Preprocessing is the first stage of the architecture, which brings the inputs image and partial caption into a normalized form that can be effectively dealt with by the systems (algorithms). The preprocessing is done on images and partial captions. In the current work, image preprocessing has been limited to image resizing and normalization. The essential steps in text preprocessing are vocabulary creation, word-to-index mapping, normalizing the length of the partial caption, word embedding vector creation, and one-hot encoding of output.

The Text encoding phase creates a hidden representation of the word embedding vector of a partial caption. Bidirectional LSTM is used for the text encoding phase.

The feature extraction phase recognizes the essential features from images needed for caption generation. The feature extraction is done in two ways: high-level feature extraction or keywords extraction and CNN feature extraction. From the experiments, CNN features are more suited for caption generation. Earlier feature extraction is done using local feature extraction techniques such as SIFT, SURF, etc., and global feature extraction techniques such as GIST, histogram, etc. Previous feature extraction techniques are time-consuming and not suited for the caption generation process. Deep learning-based feature extraction techniques give better performance on various tasks such as object classification, object detection, scene classification, etc. Various deep neural networks such as VGG, Residual Neural Network, and Densenet were experimented on the feature extraction process to determine the efficiency of the automatic description generation. The deep neural network with maximum performance is used in

the feature extraction process of the proposed methodology.

Visual attention is the process of finding a relevant part of the image suitable for the caption generation experimented. It is done on CNN features of images. A combination of spatial, channel-wise, and layer-wise attention was applied to improve the system's performance.

The caption generation phase produces the next word from the previously generated words in the description. Sequential models in deep learning are used in this phase. LSTM and Bidirectional LSTM have experimented with this process.

The optimization phase selects the good captions from the generated words. Beam search and game theoretic search are implemented for this process. Game-theoretic search outperformed beam search in description generation.

The evaluation model computes the system's performance using different evaluation metrics such as Accuracy, Loss, BLEU score, and ROUGE.

4. Visual Attention

The visual attention model is based on a multi-attention system. The multi-attention module is made up of spatial attention, channel-wise attention, and layer-wise attention. The attention network used input as the feature maps from the second last layer of Densenet with a size of $7 \times 7 \times 2208$. An attention map and a score are the outcomes of the network. The captioning module takes the attention map and captions the image using an attentive region.

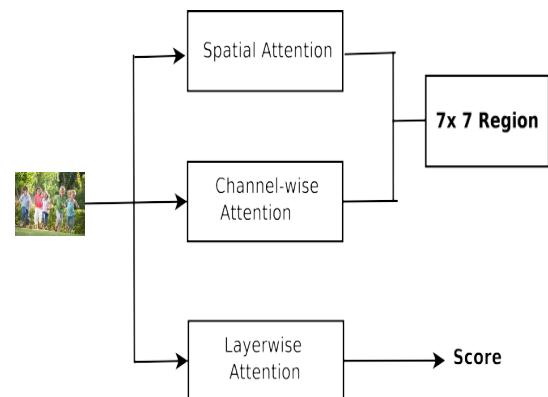


Figure 2: Visual Attention Architecture

The proposed architecture of the visual attention network is depicted in figure 2. Densenet is the convolutional neural network used for extracting the interest points of the whole image. The convolutional feature map is given to the attention network for getting a selected region and layer-wise scores. Different attention mechanisms used in the network are explained below.

4.1. Spatial Attention

The usual image captioning systems use the global feature for generating descriptions. So, it is challenging to generate a correct caption for the image based on its regions. Only local regions are taken to get an accurate description. Some regions are more peculiar than other regions in an image. The critical regions in the picture are mainly helpful in producing better descriptions. In

spatial attention, more weights are given to the necessary region despite assigning equal weights for all regions in the image.

The network of spatial attention is shown in figure 3.

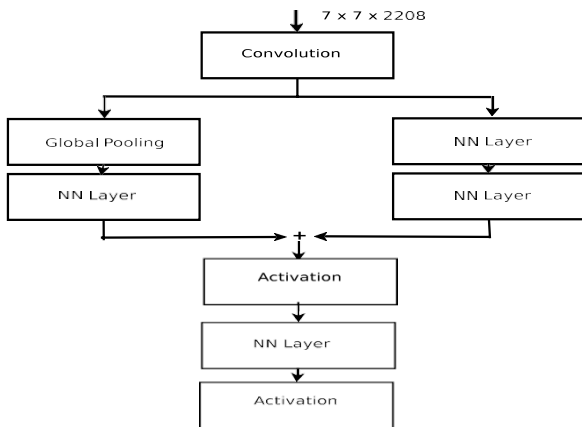


Figure 3: Spatial Attention

Let an image feature be $F \in RW \times H \times C$, F is flattened as $F_s = [f_1, f_2, \dots, f_n]$, where $f_i \in RC$, $n = W * H$ and f_i represents a spatial region i through a vector f_i . The attention weight is calculated using a single layer fully connected neural network and a Softmax function over the $W * H$ regions

$$S_a = \tanh((w_1 F_s + b_1) + (w_2 * F + b_2)) \quad (1)$$

$$S_w = \text{Softmax}(w_3 S_a + b_3) \quad (2)$$

where w_1, w_2, w_3 are weight vectors and b_1, b_2 and b_3 are the bias values for the model.

4.2. Channel-wise Attention

Colors and patterns are identified using CNN kernel functions. Some kernel functions are used to detect color information, and others are used for detecting the edges of the objects in the image. Channel-wise attention is a mechanism for choosing the channels dynamically; each channel of CNN features is obtained using the corresponding convolution kernel. The process of channel-wise attention is explained below.

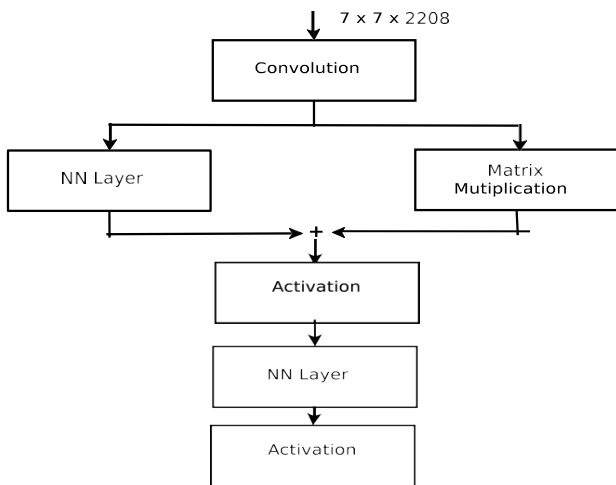


Figure 4: Channel-wise Attention

In this architecture, the image feature $F \in RW \times H \times C$ is fed to the global average pooling function to produce the channel feature $F_c = [F_1', F_2', \dots, F_c']$, $F_c' \in Rc$ Where F_i' is the output of the global average pooling function on the feature of i th channel.

$$C_a = \tanh((w_1' F_c + b_1') + (w_2' * F)) \quad (3)$$

$$C_w = \text{Softmax}(w_3' C_a + b_3') \quad (4)$$

where w_1', w_2', w_3' are weight parameters and b_1' and b_3' are the bias values for the channel-wise attention model. The modelling of channel-wise attention is explained in figure 4.

4.3. Layer-wise Attention

Different types of situations are handled by deep features in various levels. Layer-wise attention working is represented in the figure 5.

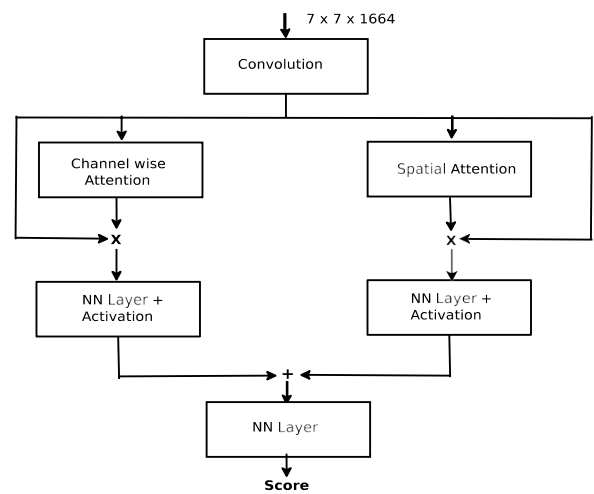


Figure 5: Layer-wise Attention

Given spatial weight S_w , channel-wise weight C_w and feature F , then

$$F_s = S_w * F \quad (5)$$

$$F_c = C_w * F \quad (6)$$

$$L_w = \text{ReLU}(w_{11} F_s + b_{11}) + \text{ReLU}(w_{21} F_c + b_{21}) \quad (7)$$

$$F_l = w_{31} L_w + b_{31} \quad (8)$$

where w_{11}, w_{21}, w_{31} are weight parameters and b_{11}, b_{21} and b_{31} are the bias values for the layer-wise attention model.

4.4. Visual Attention Parameters

The attention model starts with a convolutional layer with a kernel of size 1×1 , and the output of this layer is 512 channels. The width and height of spatial and channel-wise attention are 7. For spatial attention, the two fully connected network layers, matrix multiplication and activation functions, are integrated. The visual attention model is integrated into the caption generation system. So, the experiments are conducted for an image description generation system with visual attention.

5. Implementation Details

The framework was implemented using Keras, Tensorflow, and Python. Keras is a high-end deep learning package. The technology behind Keras is Tensorflow, which is a package for dataflow programming and machine learning.

5.1. Training Details

Image features are extracted using pre-trained Densenet model weights from ImageNet by the transfer learning mechanism. The language model used single-layer bidirectional LSTM with hidden size 256. Single-layer bidirectional LSTM, which has a hidden layer size of 1000, was employed in the caption model. The model is fitted to minimum validation loss at 50 epochs. Therefore, the model was finetuned with 50 iterations. In training, a random data generation method was employed in each iteration to limit computational resource usage. An NVIDIA Tesla K80 GPU is used for improving the training speed.

5.2. Optimization

The optimization function used was rmsprop. In rmsprop, the

learning rate for weight is divided by a running average of new gradients' magnitudes for that weight.


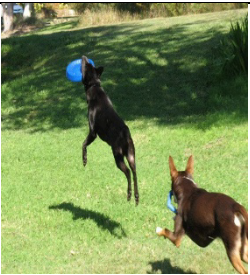
6. Experiments and Results


The performance of the model was analyzed using the BLEU[13] score. The BLEU score is an evaluation measure that compares the output description and the n-grams of the ground truth description of images. The Natural Language Toolkit (NLTK) package is used for computing the BLEU score. The experiment is done on two datasets, such as Flickr8k and ImageCLEF2019 datasets.

6.1. Flickr8k Dataset

The system is trained using a standard image captioning dataset Flickr8k[14]. The dataset is in the form of <image, caption>. The number of images in the dataset is 8000, and five captions are associated with each image. The training is done using 6000 images. The model is validated using 1000 images, and the testing is done on the remaining images. The correct image description results are shown in Table 1.



Table 1: Correct Results

Image	Ground Truth Captions	Generated Caption
	<ol style="list-style-type: none"> 1. A man crouch on a snowy peak. 2. A man in green jacket stand in a deep snow at the base of a mountain. 3. A man kneel in the snow. 4. A man measure the depth of snow. 5. A mountain hiker be dig steak into the thick snow. 	<p>Without VA+ Beam Search: A man with a stick in its mouth is standing on a snow covered field.</p> <p>Without VA +Game-theoretic search: A man with a stick is standing on a snow covered field.</p> <p>With VA+ Beam Search: A mountainer is standing in a snow covered field</p> <p>With VA +Game-theoretic search: A mountainer kneels in a snow covered field</p>
	<ol style="list-style-type: none"> 1. A dog with a Frisbee in front of a brown dog. 2. A large black dog is catching a Frisbee while a large brown dog follows shortly after. 3. Two dark colored dogs romp in the grass with blue Frisbee. 4. Two dogs are catching blue Frisbees in grass. 5. Two dogs are playing; one is catching a Frisbee. 	<p>Without VA+ Beam Search: A brown dog and a brown dog are running in a grassy field.</p> <p>Without VA +Game-theoretic search: Two brown dogs are running in a grassy field.</p> <p>With VA+ Beam Search: Two dogs are playing with a frisbee.</p> <p>With VA +Game-theoretic search: one brown dog and a dog are playing with a frisbee</p>

	<ol style="list-style-type: none"> 1. A man is sitting on the floor outside a door and his head on his chin. 2. A man sits against a yellow wall wearing all black. 3. A man wearing a dark blue hat sits on the ground and leans against a building. 4. Man with black hat, coat and pants sitting next to the door of a building. 5. The man in the black hat is sitting on the floor beside the green door. 	<p>Without VA+ Beam Search: A man in a blue jacket is sitting on a city street.</p> <p>Without VA +Game-theoretic search: A man in a blue jacket is sitting on a city street.</p> <p>With VA+ Beam Search: A man with a black hat is sitting near a door.</p> <p>With VA+ Game-theoretic search: A man with a black hat is sitting on a street near a door.</p>
---	---	---

The incorrect results are shown in Table 2.

Table 2: Incorrect results.

Image	Generated Caption
	A dog is playing with a ball
	A man with a stick in its mouth

The comparison of the model with various models is depicted in Table 3. The model was implemented with or without visual attention and caption optimization method as a beam search or game-theoretic search. Without visual attention and a game-theoretic search[15], the proposed model achieved a BLEU score of 69.96. The proposed model with visual attention and a game-theoretic search reached a BLEU score of 72.04, higher than all other models on the Flickr8k dataset given in Table 3. The results showed that the proposed model had a robust performance on the Flickr8k dataset.

Table 3: Comparison of the BLEU scores for different models.

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4
GoogleNIC [3]	63.0	41.0	27.0	-
Log bilinear[8]	65.6	42.4	27.7	17.7
Hard attention[16]	67.0	45.7	31.4	21.3
Soft attention [16]	67	44.8	29.9	19.5
Phi-LSTM [17]	67	44.8	29.9	19.5

Phi-LSTMv2 [18]	61.5	43.1	29.6	19.7
PhiLSTMv2(w.r) [18]	62.7	49.4	30.7	20.8
Our Model (W/o Visual attention + Beam search)	67.2	55.05	44.42	40.61
Our Model (W/o Visual attention + Game theoretic Search)	69.96	56.3	46.45	42.95
Our Model (With Visual attention + Beam search)	71.2	57.2	46.97	43.21
Our Model (With Visual attention + Game theoretic Search)	72.04	58.0	47.23	43.95

The proposed system is also evaluated using ROUGE score. The ROUGE score is given in the table 4. ROUGE-1, ROUGE-2 and ROUGE-L scores are computed.

Table 4: ROUGE scores of the model

Score	Precision	Recall	F-score
ROUGE-1	58.1	56.21	57.13
ROUGE-1	58.32	56.34	57.31
ROUGE-2	44.25	42.82	43.52

GCorrect is an evaluation measure for measuring the grammatical accuracy of generated descriptions. GCorrect is the average of grammatical errors in the generated captions. It is defined by Equation (9).

$$GCorrect = \sum_{i=1}^n gerror_i / n \tag{9}$$

where gerror_i is the number of grammatical errors for each

sentence and n is the number of sentences.

Grammatical errors in sentences were estimated using the Grammar-check package. The GCorrect of this framework is represented in Table 5 .

Table 5: GCorrect

Without Visual attention	0.040625
With Visual attention	0.023

6.2. ImageCLEF2019 challenge dataset[19]

The image caption pairs are extracted from PubMed Open Access. Seventy-two thousand one hundred eighty-seven radiology images are taken from the 6,031,814 image caption pairs after preprocessing. The number of images for training is 56,629 that for validation is 14,157 and for testing is 10000. Each label or symptom is mapped to a UMLS concept. The number of unique UMLS concepts is 5217. The examples of symptom UMLS concept mapping are depicted in table 6.

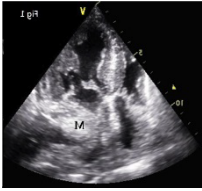
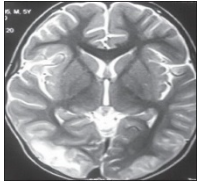

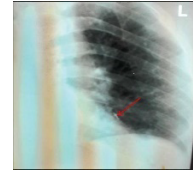
Table 6: Mapping of UMLS Concepts and Symptoms

UMLS Concept	Symptom
C0699612	osteogen
C0728713	sittings
C0376554	knowledge
C1262886	opg
C0152347	nucleus pulvinaris thalami
C0152344	capsula interna
C1550618	bronchial
C0152349	thalamus ventralis
C0520966	limb incoordination

Radiology images are resized to (224,224,3), and the intensity values of images are normalized between 0 and 1. Convolutional features of medical images are extracted using Densenet. Captions are preprocessed, and in the training set, each symptom is mapped to the concept. A vocabulary of concepts having a size 5217 is constructed. We tried our proposed image captioning model for this work.

The model is evaluated using the F1-score and mean BLEU score. Scikit-learn is used for calculating the F1-score. The default 'binary' averaging method is implemented. Some of the generated concepts with images are shown in table 7. The mean BLEU score and f1-score of the model for the medical image dataset are 25.30 and 20.56.

Table 7: ImageCLEF results

Sl. No	Image	Ground Truth Caption	Generated Caption
1		C0013516; C0203378; C0203379; C0183129; C0018792; C0221533; C0013524	C0013516; C0203378; C0203379; C0183129
2		C1552858; C0017067; C0815275; C0015252; C1258666; C0007876; C0728940; C0007776; C0022655; C0184905	C1552858; C0015252; C0007876; C0007876
3		C0043299; C1548003; C1522577; C1962945	C0043299; C1548003; C1962945
4		C0700632; C1962945; C1548003; C0179429; C0043299; C0817096; C0024109; C0796494	C1962945; C1548003; C1561542; C0043299

7. Conclusion

This paper mainly focuses on the generation of image descriptions using Deep learning methods and visual attention mechanisms. The automatic image description generation system uses encoder-decoder architecture. Different CNNs are considered for image feature extraction. Densenet gives better results for caption generation. The hybrid spatial, channel-wise, and layer-wise models are integrated into the image captioning system for producing high-quality descriptions. To optimize the words in the caption, a novel game-theoretic algorithm is introduced. Different language models are studied for generating descriptions, and

BLSTM is taken as the language model for our proposed system. An integrated framework for automatic image description generation was implemented. The model has experimented on both the general dataset Flickr8k and medical image dataset ImageCLEF2019 challenge dataset. The image and previously generated words are inputs to the integrated system. The system generated words sequentially. The system was evaluated using the BLEU score and ROUGE. The proposed method had a remarkable improvement over the state-of-the-art systems by five percentage. The grammatical correctness of the generated description was checked using a new evaluation measure called GCorrect.

References

- [1] S. Kombrink, T. Mikolov, M. Karafiát, L. Burget, "Recurrent neural network based language modeling in meeting recognition," in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2011, doi:10.21437/interpeech.2011-720.
- [2] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, **9**(8), 1997, doi:10.1162/neco.1997.9.8.1735.
- [3] O. Vinyals, A. Toshev, S. Bengio, D. Erhan, "Show and tell: A neural image caption generator," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015, doi:10.1109/CVPR.2015.7298935.
- [4] G. Kulkarni, V. Premraj, V. Ordonez, S. Dhar, S. Li, Y. Choi, A.C. Berg, T.L. Berg, "Baby talk: Understanding and generating simple image descriptions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**(12), 2013, doi:10.1109/TPAMI.2012.162.
- [5] M. Mitchell, X. Han, J. Dodge, A. Mensch, A. Goyal, A. Berg, K. Yamaguchi, T. Berg, K. Stratos, H. Daumé, "Midge: Generating image descriptions from computer vision detections," in EACL 2012 - 13th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings, 2012.
- [6] A. Karpathy, A. Joulin, F.F. Li, "Deep fragment embeddings for bidirectional image sentence mapping," in Advances in Neural Information Processing Systems, 2014.
- [7] A. Farhadi, M. Hejrati, M.A. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier, D. Forsyth, "Every picture tells a story: Generating sentences from images," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2010, doi:10.1007/978-3-642-15561-1_2.
- [8] R. Kiros, R. Zemel, R. Salakhutdinov, "Multimodal Neural Language Models," *Proc NIPS Deep Learning ...*, 2013.
- [9] Y. Gong, L. Wang, M. Hodosh, J. Hockenmaier, S. Lazebnik, "Improving image-sentence embeddings using large weakly annotated photo collections," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2014, doi:10.1007/978-3-319-10593-2_35.
- [10] A. Karpathy, L. Fei-Fei, "Deep Visual-Semantic Alignments for Generating Image Descriptions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**(4), 2017, doi:10.1109/TPAMI.2016.2598339.
- [11] M. Soh, "Learning CNN-LSTM Architectures for Image Caption Generation," *Nips*, (c), 2016.
- [12] Q. You, H. Jin, Z. Wang, C. Fang, J. Luo, "Image captioning with semantic attention," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, doi:10.1109/CVPR.2016.503.
- [13] K. R. S. W. T. Z. W. Papineni, "Bleu: A method for automatic evaluation of machine translation," in Proceedings of the 40th annual meeting on association for computational linguistics, 2002.
- [14] M. Hodosh, P. Young, J. Hockenmaier, "Framing image description as a ranking task: Data, models and evaluation metrics," *Journal of Artificial Intelligence Research*, **47**, 2013, doi:10.1613/jair.3994.
- [15] S.R. Sreela, S.M. Idicula, "Dense model for automatic image description generation with game theoretic optimization," *Information (Switzerland)*, **10**(11), 2019, doi:10.3390/info10110354.
- [16] K. Xu, J.L. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R.S. Zemel, Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in 32nd International Conference on Machine Learning, ICML 2015, 2015.
- [17] Y.H. Tan, C.S. Chan, "Phi-LSTM: A phrase-based hierarchical LSTM model for image captioning," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017, doi:10.1007/978-3-319-54193-8_7.
- [18] Y.H. Tan, C.S. Chan, "Phrase-based image caption generator with hierarchical LSTM network," *Neurocomputing*, **333**, 2019, doi:10.1016/j.neucom.2018.12.026.
- [19] A.G.S. de H. and H.M. Obioma Pelka, Christoph M. Friedrich, "Overview of the ImageCLEFmed 2019 Concept Detection Task," in CEUR Workshop Proceedings (CEUR- WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-2380/>, 2018.

Forensic Analysis of “WhatsApp” Artifacts in Android without Root

Mohammad Shadeed*, Layth Abu Arram, Majdi Owda

Department of Natural, Engineering and Technology Sciences, Arab American University Palestine, Ramallah, Palestine

ARTICLE INFO

Article history:

Received: 14 October, 2021

Accepted: 26 November, 2021

Online: 12 April, 2022

Keywords:

WhatsApp Forensics

Smart Phones Forensics

Databases Forensics

Mobile Forensics

Android Forensics

ABSTRACT

WhatsApp application is considered the largest messaging application around the world and an important source of information, they just incorporated a new technique that operates on end-to-end encryption, which presents a significant problem for forensic investigators and analysts. This study describes how to recover the encryption key from WhatsApp to decrypt WhatsApp databases and retrieve important artifacts displayed and saved in the Android system without rooting the device. As a means of presenting and analyzing artifacts taken from the most recent version of WhatsApp to signs of fresh and important evidence and artifacts to assist investigators and forensic analysts in the investigation. As a result, various techniques, devices, and software may now be utilized for WhatsApp digital forensics. The findings of this study showed a variety of artifacts in the internal memory unit utilized by the Android system for the WhatsApp application, which might aid digital forensic examiners in their examination of WhatsApp on the Android system without the need to root the device.

1. Introduction

With over one billion users, the WhatsApp instant messaging program is one of the most popular in the world. The application was founded in 2009, and Facebook bought it in February 2014 for \$ 19 billion. In exchange, Google's Android operating system is one of the most successful and well-known in smartphones throughout the world, with a market share of almost 80%, and because of the widespread use of Android devices and the services they provide, such as instant messaging, particularly the WhatsApp application, where users send and receive instant messages in their everyday lives and other activities [1], [2]. The world has well and truly entered the field of technology and its digital age, where this technology that has covered all stages and endings and forms of life is constantly presented and our main purpose is the positive use of this technology, contrary technology has facilitated our daily lives, but it has also given contributions and solutions to resist terrorist activities and electronic crimes. Specifically, this occurs as a result of technological advancements all around the world [3], [4]. According to information security experts, many crimes are committed over the Internet. Criminals may commit their crimes using a variety of channels, including the Internet, mobile devices, and instant messaging apps such as WhatsApp. With the rapid expansion of cybercrime, it has become important and urgent to begin conducting studies and research specialized in evaluating the WhatsApp program since it is regarded as the most popular application in the world that is easily utilized [5], [6].

The WhatsApp program for instant messaging is one of the most significant and popular applications in the world, with over a billion users worldwide. Because of this development and the

massive amount of information and data that is sent instantly around the world, forensic investigators and researchers have a profound and major challenge, in addition to the artifacts left inside the phone devices, which play an important role in any suspense.

This paper will provide a search for digital evidence and artifacts for the WhatsApp application, to assist digital investigation professionals and analysts in gaining access to clear scientifically proved digital evidence. Furthermore, this article will concentrate the research on artifacts from internal memory by extracting artifacts as digital evidence installed for the WhatsApp program that runs on the Android operating system. Overall, this will provide a new dimension to the digital forensic examination of the WhatsApp program.

2. Literature review

2.1. Background

This section will provide a specific and comprehensive definition of digital forensics as well as a definition of smartphones forensics. In addition, a brief description of the forensic techniques and smartphone devices data acquisition techniques used by digital forensic investigators will be provided.

2.2. Digital forensics

Digital forensics can be defined as an applied and practical use of reliable and proven methods for digital devices, and this action has been done in several ways, the most important of which are verification, identification, analysis and interpretation, and then the digital evidence that has been and derived from digital data is presented. It is the reconstruction of the events that

* Corresponding Author: Mohammad Shadeed, Email: shadeedmohammad@gmail.com

show the crime, or that helps in the anticipation of the unauthorized procedures[7]-[10]

2.3. Mobile forensics

Mobile forensics is defined as the science of digital evidence likely to be obtained from portable devices using techniques similar to digital forensic investigations. Mobile device models vary, depending on where the storage is located, so that it can be stored on the internal or external memory card, and the phone memory may be volatile or non-volatile[11]-[13].

2.4. Mobile Device Data Acquisition Techniques

Obtaining digital forensics data from mobile devices must include the use of the two main technologies, mainly logical acquisition and physical acquisition, and each of these features has its advantages [14], taking a logical copy of the device may not need to root the device and it can give us the virtual files of the stored data on the memory, while the physical version (bit by bit) needs to root the device and enables us to root the device and may cause problems on mobile devices [15]-[17].

2.5. WhatsApp forensics

Many different tools support acquisition data in the WhatsApp application, and the main goal of using these tools is to access the data stored in the protected logic of the mobile device to obtain WhatsApp data. The way these tools work depends on two methods, namely rooting the Android device and going back to the previous version of the WhatsApp application, and the use of these two methods is completely related to obtaining the data from the WhatsApp application and understanding the method on which the structure of the WhatsApp application depends. This section will discuss the WhatsApp application, and then the main mechanisms used to obtain data and items from the WhatsApp application will be clarified [18]. The WhatsApp application is an application that works on instant messaging and is free of charge, as it allows

users to receive and send messages, whether text, images, audio or video clips and many other files in a very easy way, as this application is available for all operating systems around the world such as (Windows, Android, iOS). Often thinking about how messages are transmitted from phones through the receiving servers, without thinking about what is the invisible mechanism that may take place in the phones, in fact when the sender presses the send button and directly sent the message and it is being stored inside a file and this file is being Stored inside both devices the sender and the receiver, The Figure (No.1) WhatsApp transmission. Shows Explain the process of sending and encrypting the sent message and transferring it to the WhatsApp server and then sending it to the receiver, decoding it and displaying it to the user [19]-[21].

Usually, forensic investigators are interested in locating the places where messages are stored, so they can obtain them and then run them according to WhatsApp policy, so all sent and received messages are stored in the servers of the WhatsApp application temporarily. WhatsApp servers contain a large number of evidence usually stored in servers for a very short period. In addition, investigators need to refer to the owner company directly and adhere to the company's privacy policy to retrieve any evidence in their servers, and that is a very difficult procedure. Here comes the importance of investigating the sender and receiver's devices because the sent and received messages are kept in mobile devices, so that WhatsApp users stores many artifacts of high value as proof for the investigation, as the files that are saved are the message log and the database For all conversations and correspondences (sent and received), The Figure (Figure 2) WhatsApp databases structure. shows the database structures stored in the internal memory of the mobile device hierarchically. As shown in that figure, on three levels, the first level contains the main WhatsApp data hall. The second level contains the sub-databases, and the third level, located at the bottom level of the hierarchy, shows the other sub-files [18].

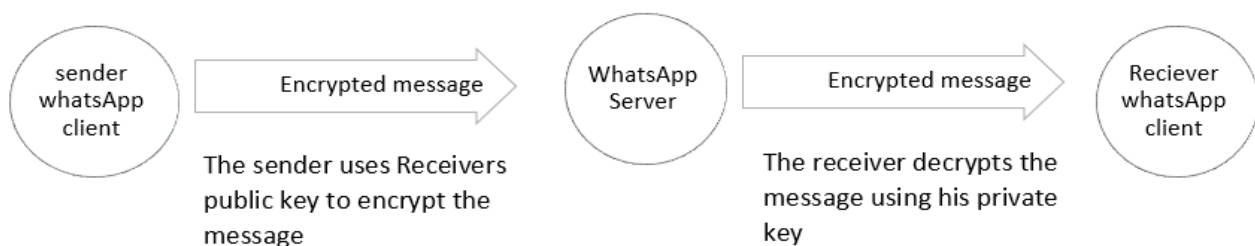


Figure 1: WhatsApp transmission

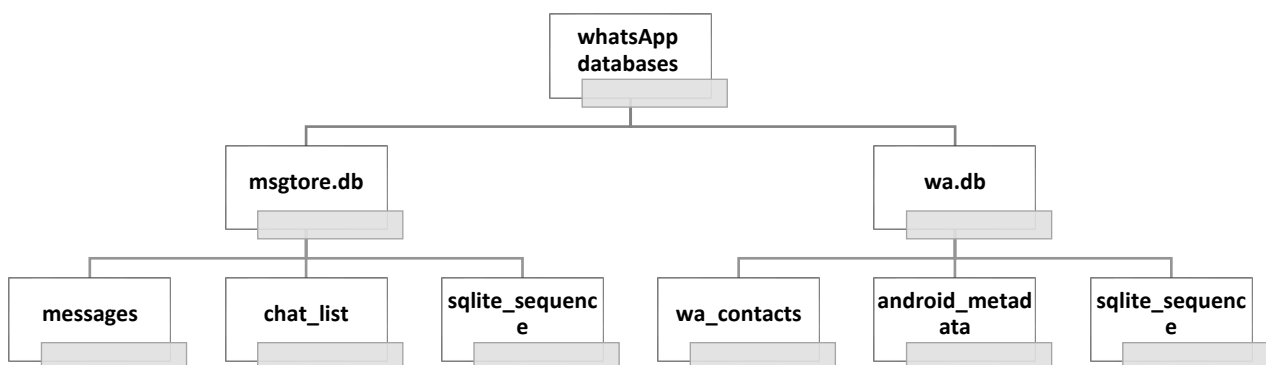


Figure 2: WhatsApp databases structure

3. Related works

There are some studies and researches that indicated this topic as a part of their efforts were unable to access or obtain sufficient antiquities as definitive digital evidence, Others have reached a portion of the artifacts. The following are some of the previous studies that were found and relevant to this research.

indicates that the data stored in the memory includes lines with characters that cannot be read by humans because it contains binary data for the full object storage array of the origin inside the database. However, the records that are observed are valuable for forensic investigation, the indexed database for WhatsApp contains with it a set of important elements that can be used as evidence of a crime after it has been presented and analyzed in the style of the time frame, and the database in WhatsApp is a valuable source of information that is used for forensic investigations and therefore the database is very important to know and see what is inside and analyze it to reach important results that may be based on the case [22]. indicate that the evidence was obtained through backup without rooting the device, and it was stored in another storage unit, and the analysis was started by open-source forensic tools, and the results showed that part of the conversations was encrypted and there were challenges in showing the results. It is not encrypted and the deleted conversations did not appear in the results, and the results indicate that to recover the deleted conversation, the device must be rooted [23]. indicates that a backup copy of Android was obtained and saved on an external memory card and analyzed on the open-source forensic analysis tools and that the database of correspondence and conversations was obtained, and even if the results obtained from Android are encrypted, the elements are evidence of correspondence, and here this study indicates that the device has not been rooted [24]. mentioned that the data of the WhatsApp conversations can be depicted in the form of a table in the memory and the chronology of the application is placed. Identification of it, and it is also possible to refer to all relevant data and what is the period for it, and this is considered as evidence or proof of my investigation for the case. Also, this method added the nature of exploration and understanding of the data in determining the time of the crime [25]. clarified that digital evidence can provide any information that has great

inferential value and showed how to interpret these data that were stored in contacts and databases of conversations and chat, and the study showed the correlation between this evidence and its connection with others to collect sufficient information that can be deduced during the examination and relates it to the event itself. Once collected, it allows the investigator to determine which messages have been deleted and remain in the log wallet in the database and log files [26]. Another research study was able to get the main supporting data from the database that contains the contacts, artifacts, and conversations that make up the WhatsApp application, and the data is in the form of a backup database and it contains relevant driver files such as images, chats, audio and video, and was able to analyze it using the applications and special tools that support to achieve the goal of the study, and that was using free tools used in digital forensics, namely (FTK) and (SQLite) browser [27]. conducted a forensic analysis process in the WhatsApp application and obtained evidence from the Android operating system on mobile devices, and the data was extracted using Python software [28]. shows the decryption of the encrypted WhatsApp database that is used in Android devices without rooting the device, and the required results were extracted, but this method does not extract the

deleted messages, but after rooting the device and using the same method, the deleted messages can be retrieved [29]. carried out a forensic analysis in WhatsApp application, where the chat and stored conversations were extracted through the internal and external memory using WhatsApp key extractor, where the decryption process was carried out and the backup database was converted into a text database so that it can be seen in the SQLite database. was based on researching the behavior of tools that use forensic evaluation in the form of extracting artifacts in the form of messages, images, or videos from Android devices. The tools (ADB WhatsApp Key / DB Extractor 4.7 and Belkasoft) were used, and all of these tools were free versions, Where the tools were tested, the WhatsApp database was extracted and decrypted, and the process of updating it using end-to-end encryption (crypt12) These results will be compared with the results obtained in this study [30].

4. Research Problem

With the ever-increasing popularity of smartphones and people's reliance on instant messages in their everyday lives, quick updates of instant messages increase the features of the application and entice users to continue using their product. On the other hand, the majority of these characteristics will provide a significant challenge to digital forensic practitioners and specialists. Many studies have been done to acquire data from older versions of the WhatsApp program. Many security measures arose with the introduction of new versions of WhatsApp, making it difficult for mobile forensic practitioners to gather information and evidence that live in internal storage. This study proposed new techniques for obtaining data from the WhatsApp application on the Android system without rooting on the latest WhatsApp application in which encrypted datasets used the new approach i.e. using crypt12 architecture.

5. Research Objectives

In this study, we will collect and analyze artifacts from the most recent release of the WhatsApp program operating on an Android machine. This study uses a variety of approaches and instruments to achieve the following goals:

- The extraction of logical WhatsApp's encrypted database, which is operating on an Android smartphone.
- Analyze and correlate the artifacts generated by the database to generate additional useful evidential trails that aid in the investigation without the usage of rooting the device.

6. Research Question

What approaches, methodologies and tools can be used in android forensics to recover artifacts from the WhatsApp Instant Message program operating on an android smartphone without rooting?

7. Methodology

The main purpose of this study is to search and find new techniques, tools, and methods to recover the artifacts of recent versions of the WhatsApp application that works in the Android operating system environment, and to extract the artifacts located in WhatsApp from the internal memory of the device. Various devices, tools, equipment, software are free and open-source and focused mainly on restoring all the artifacts that can be obtained from the latest versions of the WhatsApp application installed in the Android operating environment.

As the digital artifacts finding tools were taken through the process of logical exploration of the Android operating system. figure (NO.3) shows the approach through which the data was acquired, as it checks whether the device is rooted or not, and if it is rooted, a physical copy is taken, but if it is not rooted, a logical copy is acquired and the last stage is access to the database. And shows how the artifacts were acquired and analyzed, which were created using the WhatsApp application on smart mobile phones, in both physical and logical ways [31].

8. Requirements and Experimental Setup and Analysis

The process of forensic investigation of mobile devices is not very different from that of laptops, but some of the tools that are used in mobile forensics are somewhat different, as most mobile operating systems are closed, so it becomes difficult to understand the file system and the structure of phones. There are many open-source Android operating systems and there are some tools that are used in digital forensic for Android operating systems, which are available for users and are paid and unpaid tools, before starting the process of data extraction, backup operations, etc.

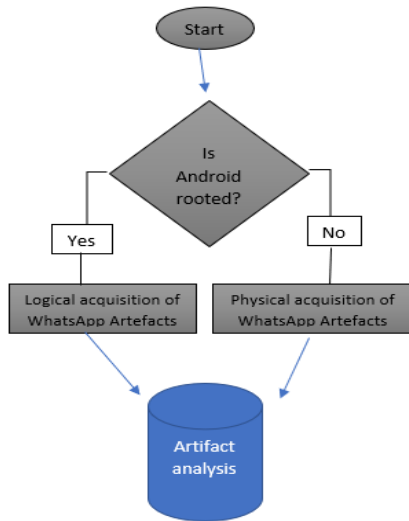


Figure 3: Methodology Design of Acquiring and Analysis of WhatsApp Artefacts

8.1. Requirements and Research tools & Devices

Table No. (I) will show the tools, software, and devices that were used in this research.

Table 1: Research tools & Devices

No .	Tools & devices	Info.
1	Redmi Note 8 (4GB RAM, 64GB) OS Android 11	Smartphone mobile
2	USB Cable	USB connector to connect phone device and computer
3	Belkasoft Evidence	Analysis tool
4	WhatsApp ver. 2.21.12.21	Messaging application
5	SQLite Studio	Analysis tool
6	SQLite Database Recovery v1.2	Tool for Database Recovery
7	Andriller	Analysis tool
9	Root Explorer 3.8	Analysis tool

No .	Tools & devices	Info.
10	Autopsy 4.4	Analysis tool
11	FINALMobile Forensics4	Analysis tool
12	WhatsApp viewer	Db viewer
13	DB Extractor	Database extractor
14	Dback	Backup viewer

8.2. Logical copy acquisition and analysis

Usually, information is stored in Android phones in different formats and ways, as a kind of security and confidentiality, so we will be careful not to change anything unnecessary on the device, it requires obtaining a logical copy by connecting the phone to the computer via a USB cable directly, from Through several tools and applications, the most important of which is to install the following tools to avoid rooting the mobile device (android-ADB, ADB fastboot, java JDK, python), and then the researcher will do an in-depth analysis of the version to obtain the (WhatsApp Artifacts) files.

This research is concerned with searching for WhatsApp artifacts in the permanent storage memory without expanding to the volatile random memory since in this research experiments were conducted on a (Xiaomi) mobile device. Application stores the data of the user (the target of this study) in a database (SQLite) called the database (msgstore.db).

After making the backup, we will find the following files

- / sdcard/WhatsApp/Databases
- / sdcard/WhatsApp/media
- / sdcard/WhatsApp/Backups

After we were able to get WhatsApp database files from the Android backup without rooting the device as shown in Figure (4), in this section of the study we will start the decryption procedures (msgstore.db) so that we can get the artifacts .

msgstore.db.crypt12	7/1/2021 08:15 ص	CRYPT12 File	291,031 KB
msgstore-2021-06-30.1.db.crypt12	6/29/2021 02:02 ص	CRYPT12 File	288,804 KB
msgstore-2021-07-01.1.db.crypt12	7/1/2021 02:02 ص	CRYPT12 File	290,994 KB

Figure 3: Whatsapp database

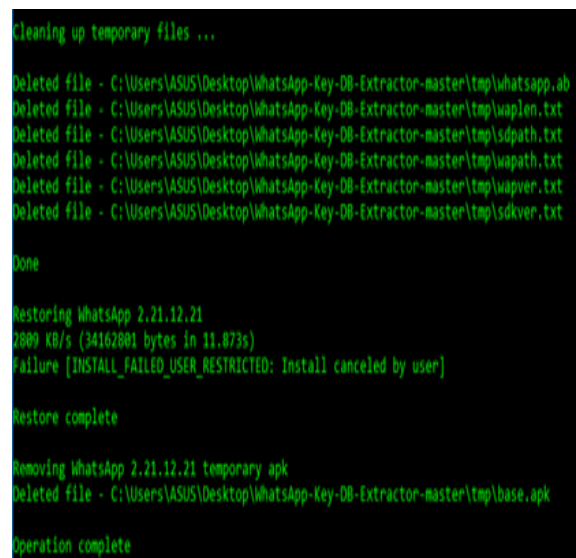


Figure 4:Extracting database

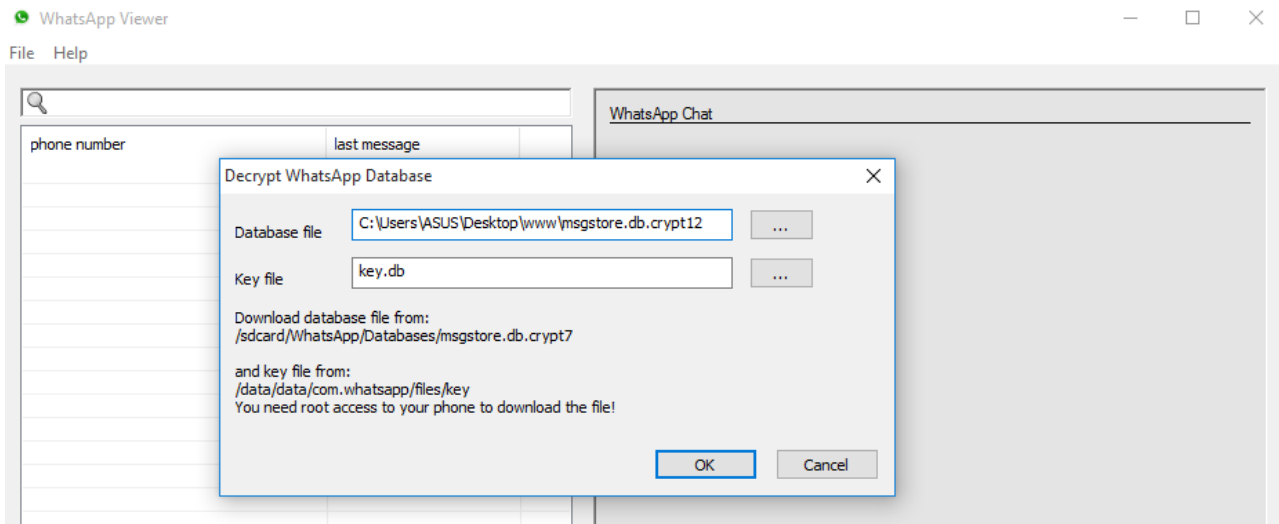


Figure 5 : To decrypt WhatsApp databases

Figure (5) shows extracting and retrieving a database opener (msgstore.db) using the specialized tool (DB Extractor) to decrypt and display WhatsApp databases. And after we were able to get the key of the database (msgstore.db) we need to decrypt the database with the same key through the (WhatsApp viewer) tool as shown in Figure (6), we will open it in order to be able to view its contents through the same tool as shown in Figure (7).

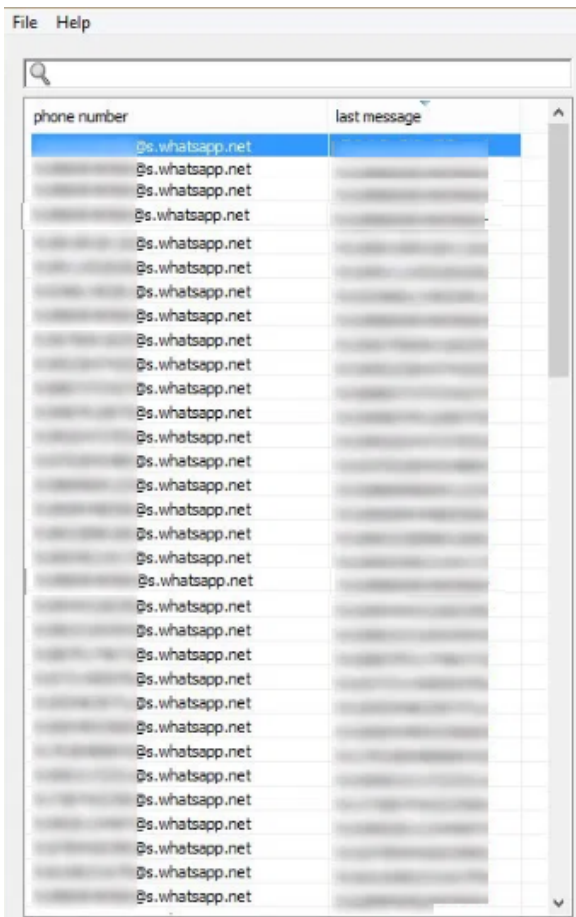


Figure 6: Open WhatsApp Database

9. Conclusion and Future Work

In this research we focused on obtaining artifacts from the device's fixed memory for the WhatsApp application, which has become well known to more than a billion users through which

an individual can communicate in all formal and informal matters by sending texts, images, documents, videos, and others. This research shows that a person can access to all the materials in WhatsApp and use other social networking applications such as Skype, Viber and Telegram that work in the operating environments of mobile devices.

All the results of this research are important and valuable results for digital forensics that can be extracted directly from smartphones that works on Android system.

As it also has been mentioned previous research's, the password will not be a black box for the device, so for a specialist to crack or overstep the password to get all the valuable user information in addition to that, files can be obtained from volatile memory (RAM) [32], [33].

Although message ,text ,user phone number and time/date of transmission are considered an important and great service provided by this study for digital forensics, but there are other matters are no less important than this, such as the deleted messages and the geographical location at the time of sending the message.

For future works. researchers will do additional researches to retrieve the deleted data in addition to the random access memory data that humans can read. There are still some experiments that can be analyzed through the volatile memory (RAM) of the device in order to obtain more artifacts that considered useful for digital forensics.

References

- [1] A.A. Ahmed, A.. Al-Qadhi, M.. Janardhana, "Geotechnical Characterization of The Volcaniclastic Rocks in and around Taiz City, Yemen," Global Journal of Advanced Engineering Technologies and Sciences, 3(4), 14–31, 2016.
- [2] E. Casey, M. Bann, J. Doyle, "Introduction to Windows Mobile Forensics," Digital Investigation, 6(3–4), 136–146, 2010, doi:10.1016/j.diin.2010.01.004.
- [3] R. Broadhurst, Y.-C. Chang, "Cybercrime in Asia: Trends and Challenges," SSRN Electronic Journal, 1–26, 2012, doi:10.2139/ssrn.2118322.
- [4] J. Liu, M. Travers, L.Y.C. Chang, "Comparative Criminology in Asia," Comparative Criminology in Asia, (October), 2–3, 2017, doi:10.1007/978-3-319-54942-2.
- [5] R. Sarre, L.Y.C. Lau, L.Y.C. Chang, "Responding to cybercrime: current trends," Police Practice and Research, 19(6), 515–518, 2018, doi:10.1080/15614263.2018.1507888.
- [6] H. Osborn Quarshie, A. Martin- Odoom, "Fighting Cybercrime in Africa,"

- Computer Science and Engineering, **2**(6), 98–100, 2012, doi:10.5923/j.computer.20120206.03.
- [7] N.V. Vukadinovic, WhatsApp Forensics: Locating Artifacts in Web and Desktop Clients, Master's Thesis, Purdue University Graduate School, 2019.
- [8] N. Beebe, "Digital forensic research: The good, the bad and the unaddressed," *IFIP Advances in Information and Communication Technology*, **306**, 17–36, 2009, doi:10.1007/978-3-642-04155-6_2.
- [9] TraceGen: User Activity Emulation for Digital Forensic Test Image Generation, Jan. 2022.
- [10] S. Omeleze, H.S. Venter, "Testing the harmonised digital forensic investigation process model-using an Android mobile phone," 2013 Information Security for South Africa - Proceedings of the ISSA 2013 Conference, 2013, doi:10.1109/ISSA.2013.6641063.
- [11] R. Singh, An Overview of Android Operating System and Its Security Features, *Engineering Research and Applications*, **4**(2), 519–521, 2014.
- [12] A. Al-Dhaqm, S.A. Razak, R.A. Ikuesan, V.R. Kebande, K. Siddique, A review of mobile forensic investigation process models, *IEEE Access*, **8**, 173359–173375, 2020, doi:10.1109/ACCESS.2020.3014615.
- [13] K. Kumar, "A Discourse of Tools for Mobile Forensic Investigation," *Researchgate.Net*, (March), 1–7, 2020.
- [14] N. Anwar, I. Riadi, "Analisis Investigasi Forensik WhatsApp Messenger Smartphone Terhadap WhatsApp Berbasis Web," *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, **3**(1), 1, 2017, doi:10.26555/jiteki.v3i1.6643.
- [15] M. Saifizi, W. Azani Mustafa, N. Syahirah Mohammad Radzi, M. Aminudin Jamlos, S. Zulkarnain Syed Idrus, "UAV Based Image Acquisition Data for 3D Model Application," *IOP Conference Series: Materials Science and Engineering*, **917**(1), 2020, doi:10.1088/1757-899X/917/1/012074.
- [16] L.H. Lee, Y. Zhu, Y.P. Yau, T. Braud, X. Su, P. Hui, "One-thumb Text Acquisition on Force-assisted Miniature Interfaces for Mobile Headsets," in 18th Annual IEEE International Conference on Pervasive Computing and Communications, *PerCom 2020*, 2020, doi:10.1109/PerCom45495.2020.9127378.
- [17] V. Arista Yuliani, I. Riadi, "Forensic Analysis WhatsApp Mobile Application On Android-Based Smartphones Using National Institute of Standard and Technology (NIST) Framework," *International Journal of Cyber-Security and Digital Forensics*, **8**(3), 223–231, 2019, doi:10.17781/p002615.
- [18] K. Alissa, N.A. Almubairik, L. Alsaleem, D. Alotaibi, M. Aldakheel, S. Alqhtani, N. Saqib, S. Brahim, M. Alshahrani, "A comparative study of WhatsApp forensics tools," *SN Applied Sciences*, **1**(11), 2019, doi:10.1007/s42452-019-1312-8.
- [19] D. Wijnberg, N.A. Le-Khac, "Identifying interception possibilities for WhatsApp communication," *Forensic Science International: Digital Investigation*, **38**, 301132, 2021, doi:10.1016/j.fsidi.2021.301132.
- [20] I.C. pada W.M.D. Forensics, Hasil cek24_60010313, 2020.
- [21] T. Sutikno, L. Handayani, D. Stiawan, M.A. Riyadi, I.M.I. Subroto, WhatsApp, viber and telegram: Which is the best for instant messaging? WhatsApp, viber and telegram: Which is the best for instant messaging?, *International Journal of Electrical and Computer Engineering*, **6**(3), 909–914, 2016, doi:10.11591/ijece.v6i3.10271.
- [22] F. Paligu, C. Varol, "Browser forensic investigations of whatsapp web utilizing indexeddb persistent storage," *Future Internet*, **12**(11), 1–17, 2020, doi:10.3390/fi12110184.
- [23] M. Iqbal, I. Riadi, "Forensic WhatsApp based Android using National Institute of Standard Technology (NIST) Method," *International Journal of Computer Applications*, **177**(8), 1–7, 2019, doi:10.5120/ijca2019919443.
- [24] J.K. Alhassan, B. Abubakar, M. Olalere, M. Abdulhamid, S. Ahmad, "Forensic Acquisition of Data from a Crypt 12 Encrypted Database of Whatsapp," 2nd International Engineering Conference, (October), 2017.
- [25] H. Shidek, N. Cahyani, A.A. Wardana, "WhatsApp Chat Visualizer: A Visualization of WhatsApp Messenger's Artifact Using the Timeline Method," *International Journal on Information and Communication Technology (IJoICT)*, **6**(1), 1, 2020, doi:10.21108/ijoiict.2020.61.489.
- [26] S. Adwan, F. Salamah, Z. Akbar, I. Krisnadi, J.K. Alhassan, B. Abubakar, M. Olalere, M. Abdulhamid, S. Ahmad, K. Alissa, N.A. Almubairik, L. Alsaleem, D. Alotaibi, M. Aldakheel, S. Alqhtani, N. Saqib, S. Brahim, M. Alshahrani, C. Anglano, D.A.O. and A. Castro2, Fitria, H.A. Ghannam, A. Hamid, F. Ahmad, K. Ram, A. Khalique, M. Mirza, F.E. Salamh, U. Karabiyik, et al., Forensic analysis of whatsapp messenger on Android smartphones, *International Journal on Information and Communication Technology (IJoICT)*, **6**(1), 1–17, 2020.
- [27] B. Actoriano, I. Riadi, "Forensic Investigation on Whatsapp Web Using Framework Integrated Digital Forensic Investigation Framework Version 2," *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, **7**(4), 410–419, 2018.
- [28] M.S. Sahu, "An Analysis of WhatsApp Forensics in Android Smartphones," *International Journal of Engineering Research*, **3**(5), 349–350, 2014, doi:10.17950/ijer/v3s5/514.
- [29] K. Rathi, U. Karabiyik, T. Aderibigbe, H. Chi, "Forensic analysis of encrypted instant messaging applications on Android," 6th International Symposium on Digital Forensic and Security, ISDFS 2018 - Proceeding, 2018-January, 1–6, 2018, doi:10.1109/ISDFS.2018.8355344.
- [30] R. Umar, I. Riadi, G.M. Zamroni, "Mobile forensic tools evaluation for digital crime investigation," *International Journal on Advanced Science, Engineering and Information Technology*, **8**(3), 949–955, 2018, doi:10.18517/ijaseit.8.3.3591.
- [31] H.A. Ghannam, "Forensic Analysis of Artifacts of Giant Instant Messaging 'WhatsApp' in Android Smartphone," *Journal of Applied Information, Communication and Technology*, **5**(2), 63–72, 2018, doi:10.33555/ejaict.v5i2.55.
- [32] O. Wee Sern, N. Hidayah Ab Rahman, F. Sains Komputer dan Teknologi Maklumat, U. Tun Hussein Onn Malaysia, P. Raja, B. Pahat, "A Forensic Analysis Visualization Tool for Mobile Instant Messaging Apps," *Intl. Journal on ICT*, **6**(2), 78–87, 2020, doi:10.21108/IJOICT.2020.00.530.
- [33] R.D. Thantilage, N.A. Le Khac, Framework for the retrieval of social media and instant messaging evidence from volatile memory, *Proceedings - 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering, TrustCom/BigDataSE 2019*, 476–482, 2019, doi:10.1109/TrustCom/BigDataSE.2019.00070.

An Interdisciplinary Approach to Fracture of Solids from the Standpoint of Condensed Matter Physics

Mark Petrov*

Department of Strength and Durability of Materials and Structural Components, Aeronautical Research Institute named after S. A. Chaplygin, Novosibirsk, 630051, Russia

ARTICLE INFO

Article history:

Received: 27 January, 2022

Accepted: 14 March, 2022

Online: 12 April, 2022

Keywords:

Strength

Creep

Fatigue

Inelasticity

Rheology

Damages

ABSTRACT

Instead of approaches of solid mechanics or a formal description of experimental data an interdisciplinary approach is proposed to consider failure and deformation as thermodynamic processes. Mathematical modeling of the processes is carried out using rheological models of the material. One fracture criterion is used, that formally corresponds to the achievement of a threshold concentration of micro-damage in any volume of the material. The prediction of the durability of materials under constant or variable temperature and force conditions is performed by time steps, including situations with changes in the material structure. Calculations of durability of structural components are based on the relationship of plastic flow and failure processes distributed over the volume of the material.

1. Introduction

In our article we have shown the possibilities and necessity of applying an interdisciplinary approach to solving the problem of flow and fracture of materials [1]. Kauzmann was one of the first scientists who applied the theory of reaction rates to the yielding of solids, examining creep as a process of directional diffusion under the effect of applied stresses [2]. Assuming that the applied stress reduces the energy barrier in one direction and increases the barrier to approximately the same extent in the opposite direction, he derived an equation for the excess number of transfer acts per unit time in the direction of applied stresses:

$$n_* = 2A \exp\left(-\frac{H}{kT}\right) \sinh\left(\frac{\Delta H}{kT}\right), \quad (1)$$

where H is the initial height of the energy barrier, ΔH is the variation of this height under the stress effect, k is the Boltzmann constant, T is the absolute temperature, and A is the reaction constant. At high values of ΔH , the reverse flow through the barrier is usually ignored, and Eq. (1) takes the form

$$n_* = A \exp\left(-\frac{H - \Delta H}{kT}\right). \quad (2)$$

Kauzmann assumed that ΔH depends in a linear manner on stresses, and (2) was confirmed by some experiments.

If some process in the material is caused by thermal activation, the dependence of the process rate on the stress σ and temperature is described by the Arrhenius equation in which the pre-exponential factor depends in the general case on the stress and temperature. Specific types of the expressions $V_0(\sigma, T)$ and activation energy $U(\sigma)$ are determined by the range of temperature–force conditions of loading. Each such region is characterized by the dominance of some deformation mechanism or mass transfer mechanism. The physical interpretation of this equation is based on the theory of overcoming potential barriers. The exponent in (3) is interpreted as the probability of the transition through the barrier or as the fraction of atoms that are in the activated state at each time instant [3].

$$V = V_0(\sigma, T) \exp\left[-\frac{U(\sigma)}{kT}\right], \quad (3)$$

Based on this concept, the analysis of strength and deformation characteristics of any material should be started from the analysis of results of simple experiments on fracture at constant stress and temperature to identify the basic features of these processes corresponding, for example, to the form (2). Tests performed under monotonic loading provide additional data, which may ensure a more accurate description of the material behavior [4, 5]. In this case, the main research method is the thermally activation analysis.

* Corresponding Author: Mark Petrov1, post box 166, 630089, Novosibirsk, Russia, markp@risp.ru

With cyclic loading with small amplitudes leading to fatigue failure, thermal activation analysis cannot be performed. The process of failure is localized and distributed in the volume of the material according to the field of internal stresses. The problem can be solved only by mathematical modeling of local processes of fracture, based on the same patterns of fracture that were revealed during the study of the creep of the material.

There are objective reasons for the lack of reliable methods of durability calculations. Because of the large variety of operation conditions, investigations were separated into individual fields, and the bearing capacity of particular structural elements was studied only for particular conditions of their operation. Internal processes in the material under fracture are rather complicated and versatile; the lack of information about their relationship with macroscopic properties of solids gave rise to many approaches both to understanding the fracture phenomenon and to developing engineering methods of estimating the bearing capacity of various structures.

Despite comprehensive investigations, there is no unified concept, which would allow successful evaluation of strength and durability of structures under hostile conditions of their operation. There are many publications dealing with physical and metal science aspects of strength and durability. These studies assist in understanding what happens in the material and explain experimentally observed specific features of the material behavior. However, such studies are not directly related to calculations of strength and durability in practice. There are many approaches and methods for determining the bearing capacity of structures depending on the loading character and temperature, though each of these approaches and methods is applicable only in a limited range of operation conditions. If the range of operation conditions is extended, there arises a problem of matching these approaches. The problem is difficult because the basis of the problem solution, i.e., the material itself, is ignored. It is sufficient to say that even different units of durability measurement are different for different loading types: these may be the time, or the number of cycles, or even the sum of loads. It is necessary to revise the traditional methods used for estimating strength and durability of structures from the viewpoint of physics of material properties.

The solution of the problem of assessing the durability of materials in structures under arbitrary thermal-force loading is impossible without constructing new models of continuous media on the basis of physics and thermodynamics of internal processes that occur in loaded solids. It is only an adequate presentation of a solid as a physical medium that offers a possibility of considering the entire multitude of interrelated processes of deformation and fracture, structural transformations, and physical and chemical effects. The analysis of experimental data from this viewpoint leads to qualitatively new ideas of material properties and allows determining the optimal volume of the experiment and the sequence of obtaining the characteristics of new alloys, thus, reducing the cost and time of structural design.

2. Basic laws of failure and deformation of materials

Examination of the kinetics of fracture of polymers, pure metals, alloys, and other materials showed that the following dependence of durability as the inverse of the average rate of failure $\dot{\omega}$ on the temperature and stress is satisfied in many cases

(for a mole of a substance by replacing the Boltzmann constant k with the universal gas constant R):

$$\tau = \tau_0 \exp\left(\frac{U_0 - \gamma\sigma}{RT}\right), \tag{4}$$

or in the general case when temperature, stress and parameter γ depend on time t ,

$$\dot{\omega} = v_0 \exp\left(-\frac{U_0 - \gamma(t, \sigma, T)\sigma(t)}{RT(t)}\right), \tag{5}$$

where U_0 is initial activation energy of fracture, γ is the structure-sensitive coefficient (activation volume), $v_0 = 1/\tau_0$ – characteristic Debye frequency [6, 7]. The expression for the plastic strain rate at a constant stress (steady creep stage) obtained in the same experiments has a similar form

$$\dot{\epsilon}_p = \dot{\epsilon}_0 \exp\left(-\frac{Q_0 - \alpha\sigma}{RT}\right), \tag{6}$$

indicating a close relationship of the fracture processes with the processes of plastic deformation. A comparison of the parameters of (4), (5) and (6) for many materials in fact shows the equality (within the limits of the error of experimental data processing) of U_0 and Q_0 , γ and α , and the product $\tau_0\dot{\epsilon}_0$ is equal to the residual strain ϵ_s accumulated at the steady creep stage [6]. The residual strain changes only slightly (approximately by an order of magnitude) with a large change in the duration of fracture (9–10 orders) [7]. The values of the pre-exponential factors in (4) and (6) determined in processing of experimental data for different materials were in the range 10^{-11} – 10^{-14} s for τ_0 and 10^{12} – 10^{13} s⁻¹ for $\dot{\epsilon}_0$.

In many cases, the activation characteristics of atomic rearrangement processes are reflected in the macroscopic characteristics of the solid under loading. Therefore, natural attempts have been made to examine the mechanism of these processes by means of the thermal activation analysis. For this purpose, in accordance with (2), (3), (4) or (5), we plot the logarithm of the process rate on the stress at different temperatures and on the reciprocal value of temperature and different stresses.

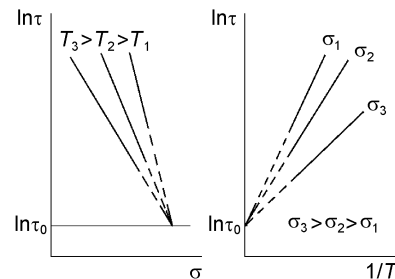


Figure 1: Temperature–force dependences of durability for determining the activation parameters of the fracture process

If, for example, (4) is valid and its parameters are constant, then we obtain a series of straight lines in the corresponding coordinates, with the lines converging in a band (Fig. 1 and 2). At the same time, it is evident that the process itself may lead to changes in the state

of the medium in which it takes place. This results in changes of the parameters and in their dependence on both the external conditions (σ , T) and the stage of the process, i.e., internal conditions. For materials, these are structural changes, being the result of the combined effect of different atomic mechanisms in different configurations at each scale level.

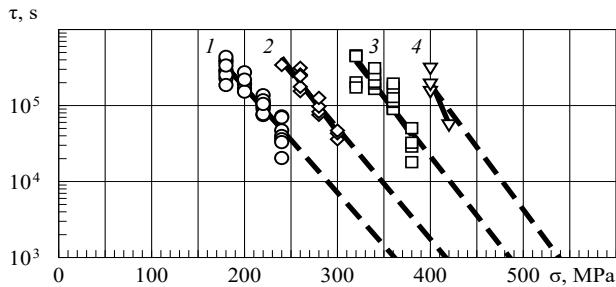


Figure 2: Temperature–force dependences of durability for determining the activation parameters of the fracture process

In the article [1], the force dependences of the activation energy of fracture and deformation for alloy 1201 T1 were shown in Figure 1 (Al-Cu-Mn system). The same refers to duralumin (the durability of its specimens is shown in Figure 2, Al-Cu-Mg system). The force dependence of the activation energy of fracture obtained in this experiment is shown in Figure 3. In the range of stresses and temperatures in which they were tested, the coefficient γ also has a constant (minimum) value, and the specimens demonstrate structurally stable “long-term strength,” which can be recalculated from one temperature-force mode to another. Similar dependences were previously given earlier for this and other aluminum alloys, including quasi-stable states at $\gamma = \gamma_{max}$ [1, 4, 5, 8, 9]. All this can be seen only through the thermal activation analysis, taking into account, among other things, the quantum effects of low-temperature fracture of materials [9].

For mechanical engineers who are used to terms “strength” or “long-term strength,” we can offer a more stringent strength characteristic – the strength parameter. Let us define it as $P_b = 1/\gamma$. Then the above-mentioned specimens in a certain temperature-time interval, regardless of the loading speed and temperature, will have $P_b = const$, i.e., exactly the same “strength” determined only by the material structure (activation volume γ). The dimension of P_b is MPa·mol/kJ, and it is also independent of the loading trajectory if the loading rate is variable. Thus, the comparison of the “strength” tests results is justified.

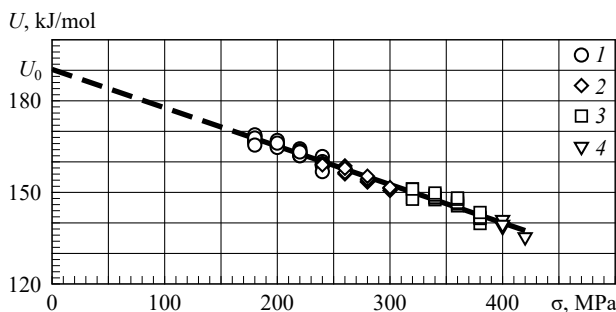


Figure 3: Force dependences of the activation energy of fracture of 80 duralumin specimens (Fig. 2) in the range of stresses of 180–420 MPa and temperatures of 398–473 K; temperature, K: 1 – 473, 2 – 448, 3 – 423, 4 – 398

For mechanical engineers who are used to terms “strength” or “long-term strength,” we can offer a more stringent strength characteristic – the strength parameter. Let us define it as $P_b = 1/\gamma$. Then the above-mentioned specimens in a certain temperature-time interval, regardless of the loading speed and temperature, will have $P_b = const$, i.e., exactly the same “strength” determined only by the material structure (activation volume γ). The dimension of P_b is MPa·mol/kJ, and it is also independent of the loading trajectory if the loading rate is variable. Thus, the comparison of the “strength” tests results is justified.

An extensive experiment analyzed by the methods described here revealed the influence of creep of the binder on the strength properties of fiberglass plastic [10]. Figure 4 shows the force dependences of the activation energy of fiberglass plastic fracture under longitudinal bending (a) and tensile loading (b). For longitudinal bending, the mean values for 20 or 40 specimens tested in each mode are shown. The diamonds denote modes of monotonic loading with different rates, and the circles show loading by a constant bending moment. For tensile loading, the data for each specimen are provided. The stress scatter corresponds to monotonic loading, and the scatter of the activation energy corresponds to a constant load.

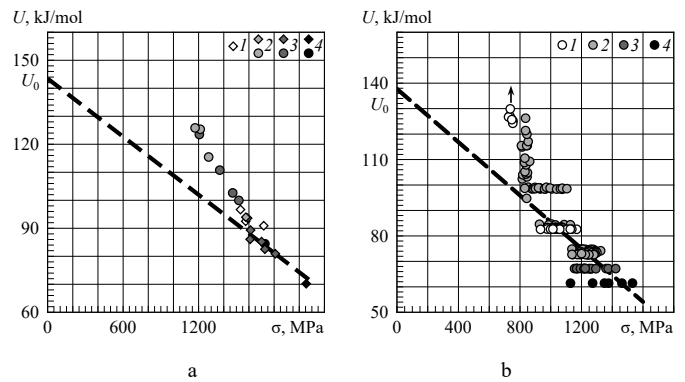


Figure 4: Force dependences of the activation energy of fiberglass fracture (rods with a diameter of 5.5 mm) under longitudinal bending (a) and tensile loading (b) [10]; temperature, T , °C: a) 1 – +60, 2 – +50, 3 – +20, 4 – –30; b) 1 – +50, 2 – +9–20, 3 – –8, 4 – –30

The lines drawn through those values of $U(\sigma)$ that satisfy the equation of a straight line with the minimum slope showed approximately the same values of the initial activation energy U_0 . The deviations from these lines illustrate the role of binder creep in the distribution of forces over the fibers in the composite, which is equivalent to changing the material structure. Thus, the straight lines correspond apparently to some stable state of the material structure, when the initial stage of creep has already ended. With a decrease in stresses and an increase in the duration of the failure process, the experimental values of $U(\sigma)$ deviate from these lines in the direction of increasing durability. Rapid loading or temperature reduction leads to more uneven loading of the fibers, and they begin to break down sequentially. The rod failure process is finalized even at lower loads [10] in contrast to metals in which the rupture stresses of the specimens increase in proportion to the growth of the logarithm of the loading rate (or to the decrease in the logarithm of the fracture time).

Under monotonous loading, the values of $U(\sigma)$ are calculated by the equivalent failure time τ_{eq} in accordance with (5) reduction

to maximum stresses, based on the same damage in accordance with the Bailey criterion [11], according to the formula

$$\tau_{eq} v_0 \exp\left(-\frac{U_0 - \gamma \sigma_{max}}{RT}\right) = \int_0^{t_*} \dot{\omega} dt = 1, \quad (7)$$

where t_* is the loading time along a particular trajectory to a stress σ_* , often less σ_{max} , at which specimen fracture occurs. The error in determining the activation volume γ turns out to be small, since the entire process of failure is short-lived and concentrated in the range of action of high stresses.

The change in the activation volume γ is associated with a change in the structure of the material and each material has its own reasons and characteristics. These can be, for example, relaxation processes of internal stresses, diffusion of alloying elements in the alloy matrix or creep of the binder in the composite material. And each of them requires separate study and modeling.

3. Mathematical modeling of the rheological properties of the material

In accordance with the patterns of fracture and deformation of materials (1) and (2) observed in the experiment, new bodies were introduced into rheology, called Zhurkov (Zh) and Kauzmann (Km) bodies [12]. Denoting $A = \varepsilon_* v_0 \exp(-Q_0 / RT)$ and $B = \alpha / RT$ in (6), we obtain the rheological equations of the Zh and the Km solid:

$$\dot{\varepsilon}_p = A \exp(B\sigma) \text{ and } \dot{\varepsilon}_p = 2A \operatorname{sh}(B\sigma). \quad (8)$$

The sequential and parallel connections of these bodies with the Hooke body (H) having an elastic modulus M form bodies similar to the Maxwell and Kelvin (Voigt) bodies, which describe the general and local plastic flow in materials. They are indicated by the symbols PM_1 and PM_2 (with Zh bodies) or PM_5 and PM_6 (with Km bodies) [1, 12]. A set of such elements is a structural model of the material that shows in Figure 5. The difference from the similar mechanical structural model of the material based on the Saint-Venant body [13] is the replacement of dry friction elements with elements that describe plastic flow kinetics (8). An element of the general plastic flow of the material has also been added.

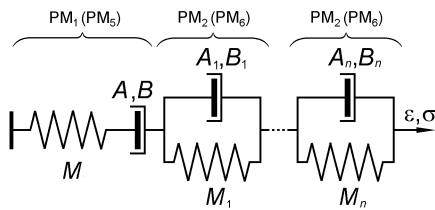


Figure 5: Structural model of a material describing the elasticity, creep, and hysteresis-type inelasticity by a set of elements with a parallel connection of an elastic Hooke's body and a plastic flow body (Zhurkov's or Kauzmann's body) [1, 4]

The plastic flow and the process of material failure are both occurring and interrelated [14]. Therefore, describing the rheological properties of materials, it is possible to associate them with the accumulation of damage and to assess the durability of structural elements under arbitrary external effects [5].

The rheological equation, for example, of the PM_1 solid (Fig. 5) as the equality of the total strain rate to the sum of the elastic strain rate of the H solid and the plastic strain of the Zh solid (8) has the following form [1, 12]:

$$\dot{\varepsilon} = \frac{\dot{\sigma}}{M} + A \exp(B\sigma). \quad (9)$$

This equation has the following solutions [12]:

- at a constant strain ($\dot{\varepsilon} = 0$), we obtain the equation of stress relaxation

$$\sigma = -\frac{1}{B} \ln[\exp(-B\sigma_0) + ABMt]; \quad (10)$$

- at a constant strain rate ($\dot{\varepsilon} = C$), substituting the integration constant in (10) by the function and deriving the linear equation, we obtain the expression

$$\sigma = -\frac{1}{B} \ln \left\{ \exp[-B(\sigma_0 + MCt)] + \frac{A}{C} [1 - \exp(-BMCt)] \right\} \quad (11)$$

As the time progresses ($t \rightarrow \infty$), this expression yields the flow stress (yield stress)

$$\sigma = -\frac{1}{B} \ln\left(\frac{A}{C}\right), \quad (12)$$

which depends on the strain rate and temperature;

- at loading with a constant rate $\dot{\sigma} = D$, we obtain the dependence of strain on time

$$\varepsilon = \varepsilon_0 + \frac{Dt}{M} + A \exp(B\sigma_0) \frac{\exp(BDt) - 1}{BD}. \quad (13)$$

At $D = 0$, this dependence transforms to the steady creep equation

$$\varepsilon = \varepsilon_0 + A \exp(B\sigma_0)t$$

Here σ_0 and ε_0 are the stress and strain at the time instant $t = 0$.

Solutions of (9) for a constant strain rate C (11) or for a constant loading rate D (13) give two different relationships between stresses and strains. As a result, we get several "theories of plasticity" [5]. If the material structure changes during plastic flow, the parameters of (9) should be replaced with functions describing the transition of the material from one state to another. This can be done by analyzing the experimental deformation curves by time steps [1, 12]. No new "theories of plasticity" are required. And stress relaxation according to solution (10), in which the material also is also fractured, does not require any energy expenditure. Everything happens due to the internal energy of a solid, the measure of which is temperature.

For the PM₂ solid (Fig. 5), on the basis of solving the equilibrium and strain compatibility equations, we can write the following rheological equation:

$$\dot{\varepsilon} \exp(BM\varepsilon) = A \exp(B\sigma). \quad (14)$$

Integration of this equation for $\sigma = \sigma_0 + Dt$ yields the solution

$$\varepsilon = \frac{1}{M} \left[\sigma_0 + Dt + \frac{1}{B} \ln \left\{ \frac{\exp[-B(\sigma_0 - M\varepsilon_0 + Dt)]}{+ \frac{AM}{D} [1 - \exp(-BDt)]} \right\} \right]. \quad (15)$$

For the loading rate $D = 0$, we obtain

$$\varepsilon = \frac{1}{M} \left[\sigma_0 + \frac{1}{B} \ln \{ \exp[-B(\sigma_0 - M\varepsilon_0)] + ABMt \} \right]. \quad (16)$$

When stress relaxation (10) or strain relaxation (16) occurs, and the stresses reach a small value, one should use similar solutions of rheological equations with Kauzmann bodies (at $\sigma \rightarrow 0$ the rate of plastic strain $\dot{\varepsilon}_p \rightarrow 0$) [12]. This should be especially borne in mind at elevated temperatures.

For a PM₅ body with constant strain, instead of (10) we have

$$\sigma = \frac{2}{B} \operatorname{artanh} \left[\tanh \left(\frac{B\sigma_0}{2} \right) \exp(-2ABMt) \right], \quad (17)$$

or, for the calculation procedures [15],

$$\sigma = \frac{1}{B} \ln \frac{1+X}{1-X}; \quad \left(X = \frac{\exp(B\sigma_0) - 1}{\exp(B\sigma_0) + 1} \exp(-2ABMt) \right).$$

For the PM₆ body at constant stresses, instead of (16) we obtain

$$\varepsilon = \frac{1}{M} \left[\sigma_0 - \frac{2}{B} \operatorname{artanh} \left\{ \tanh \left[\frac{B(\sigma_0 - M\varepsilon_0)}{2} \right] \times \exp(-2ABMt) \right\} \right], \quad (18)$$

Solutions (10), (16) and (17), (18) provide completely identical results in the range of high stresses. Therefore, when using the Km solid in the models, it is more efficient to use the solutions for models with the Zh solid in appropriate sections of the loading program, because this is a simpler procedure. This also refers to the algorithms of processing the experimental data for determining the parameters of the rheological models. Figure 6 shows the stresses in the PM₁ and PM₅ solids with their rapid deformation to the establishment of constant flow stresses and subsequent curing at a fixed strain. Calculations were carried out for duralumin: $Q_0 = 192.6$ kJ/mol, $\alpha = 0.142$ kJ/(mol·MPa).

If the material is characterized by the same behaviour in tensile and compressive loading, it is only necessary to change the signs of the stresses, strains, and their rates to the opposite signs when passing to the compression region. Otherwise, the parameters A

and B should also differ. The algorithm of calculations in the transition through zero should be also accurate. The time step in unloading should be selected in such a manner that the stresses in the flow elements should approach zero prior to “reversing” of the equations. Otherwise, the strains would be determined with errors.

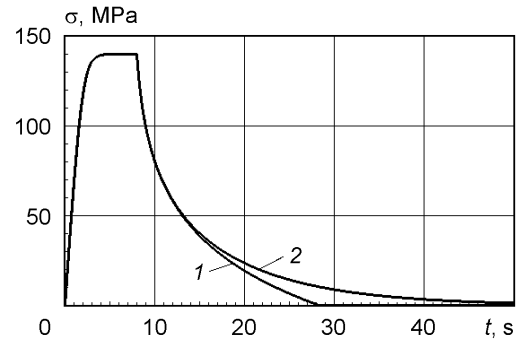


Figure 6: Deformation of the PM₁ (1) and PM₅ (2) solids with a constant strain rate to the “yield limit” and subsequent stress relaxation at a constant strain (the parameters of the rheological solids were taken for the D16 T material at 573 K)

When the stresses are greater than $-\ln(AM/D)/B$ in (15), there is something like a functional relationship between the stresses and strains in subsequent loading. In this case, we have “plasticity with hardening.” If loading is terminated, we obtain equations of the so-called logarithmic creep [16], which were interpreted analytically in (16). In processing experimental data for an actual material, it is necessary to separate the plastic flow with actual hardening accompanied by changes in the material structure and by a decrease in the activation volume α from the local flow. The latter takes place in a set of local volumes; in each volume, it is characterized by its own activation parameters.

As a test problem, we study with a stress jump. Experimental results of such tests have long been known [16]: in the unsteady stage of creep, the initial stress increased or decreased by a jump, and after some time it returned to the same level.

Figure 7 gives results of calculations of the deformation process in the D16 T alloy performed using its model. The flow characteristics of the material predicted by the model are exactly the same as those of the real flow. No additional conditions apart from specification of the loading program and the temperature are required [12].

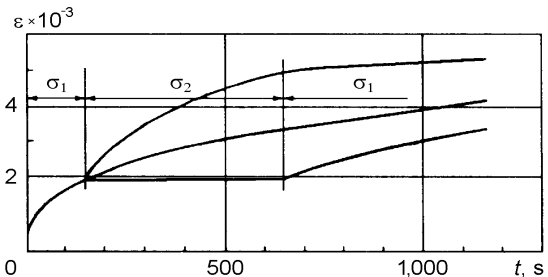


Figure 7: Creep with a stress jump: calculations using the model of the material D16 T (423 K); $\sigma_1 = 300$ MPa, $\sigma_2 = 270$ and 310 MPa

Other examples of the calculation of strains of specimens during loading and unloading are given in the article [5]. Experimental data, with which the results of calculations are compared, are contained in the article [17] or obtained by the author himself.

For such calculations based on the amplitude dependence of inelasticity, parametric identification of the structural model of the material (PM₂ or PM₆ bodies) is performed by dividing it into components that characterize each structural element. The typical amplitude dependence of the inelastic deformation of the material in the form of the width of the inelasticity loop is shown in Figure 8.

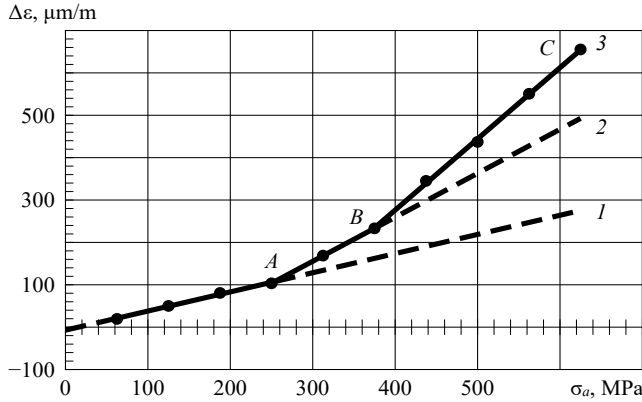


Figure 8: An example of a typical amplitude dependence of the inelasticity loop width of an eight-layer unidirectional carbon fiber reinforced plastic

The broken line 3 in the figure shows the value of the loop width: the maximum distance between the loading and unloading curve $\epsilon = f(\sigma)$, calculated at a constant mean value of the cycle stresses. The data are taken from an experiment performed on unidirectional carbon fiber reinforced plastic. Up to point A (line 1), there is always relaxation-type inelasticity in any material [18, 19]. As the loading amplitude increases, hysteresis-type inelasticity additionally appears (segment AB on line 2). This is followed by a new increase in inelasticity (segment BC). Each loop width increment is ascribed to one structural element of the material model, which will determine its durability in the corresponding range of amplitudes.

The dependence of durability on mean cyclic stresses is taken into account in the rheological model of the material by changing the loop width through the change of the parameter $\dot{\epsilon}_0$ in (6) for each structural element of the material model. For this purpose, in each amplitude range it is necessary to test with a different asymmetry index [4], and the endurance value N (the number of cycles passed during the specimen fracture) will be inversely proportional to the increment of the loop width in this range.

After parametric identification of the mathematical model carried out using experimental data for a specific frequency and temperature of tests, it is possible to proceed to calculations of the durability of the material under arbitrary changes in temperature and stress within the studied range of temperature-force dependences of the deformation activation energy and the fracture activation energy. When the material structure changes, the parameters A and B in (8) should be replaced by functions describing the accompanying thermally activated processes or the results of some other external effects leading to these changes.

An example of this is the fracture of duralumin at various combinations of temperature and stress. A precipitation aged alloy, which has reached the first maximum of hardness, undergoes a phase aging stage in the process of failure (intermetallic

precipitation). Its hardness, having reached the second maximum, begins to decrease. This also happens in the absence of stresses, and the process accelerates under load, which affects the residual strain of the specimens. Figure 9 shows the dependence of the residual strain of duralumin specimens on the tensile test mode. The observed minimum of residual strain during the period of steady creep is associated with the achievement of the maximum hardness of the material.

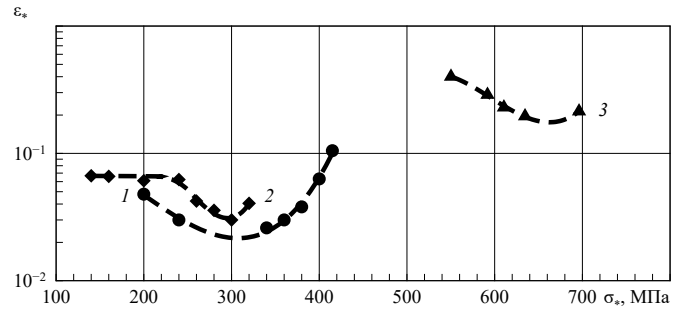


Figure 9: Dependence of the residual strain of duralumin specimens on the stresses and temperature of the tests (creep under constant and monotonically increasing loads); T, K : 1 – 423, 2 – 448, 3 – 293÷523; 1, 2 – average values for several specimens, 3 – true residual strains in the neck of specimens under monotonic loading; σ_* – initial stress values at constant load (1, 2) and the highest stress values at specimen rupture (3).

The change in hardness over time of alloys similar in composition was studied in [20–23] and others. The time to reach a certain state of the alloy in the process of transformation, which occurs due to thermal activation, is determined by a typical expression of the theory of the rates of processes [24]

$$\tau_p = \tau_{p0} \exp\left(\frac{Q_p}{RT}\right), \quad (19)$$

where Q_p is the activation energy of precipitate growth, the value of which is equal to the effective value of the activation energy of diffusion of alloying elements in the alloy matrix. The activation energy of diffusion, similar to what we observe during failure, depends approximately linearly on stresses, and the diffusion process is accelerated as a result of plastic deformation [18]. Then, taking into account the simultaneity of the process of precipitations in the centers, which additionally arise due to the accumulated plastic strain ϵ_{res} , expression (19) takes the form

$$\tau_p = \tau_{p0} \exp\left(\frac{Q_p - \beta\sigma}{RT} - m\epsilon_{res}\right). \quad (20)$$

In the case of an additional increase in the number of precipitation centers due to a greater concentration of vacancies, the pre-exponential multiplier in (20) should include a multiplier expression that takes into account the temperature of the cooling medium during quenching [18]. At this temperature, the equilibrium concentration of vacancies has time to be established [20], and in this form this expression can participate in the calculations of the aging process at low temperatures and in the description of the recovery process. The characteristic maximum hardness of isolate-aged alloys and the corresponding minimum of

plasticity make it possible to estimate the activation parameters in the expression (20).

By equating the failure time (4) and the time to reach the minimum of the residual creep strain (20) obtained under certain loading conditions (Fig. 9), it is possible to estimate the parameters included in (20). Having the parameters of (20), we obtain an expression for the conditional rate of the precipitation process $V_p = 1/\tau_p$, the integral of which in time will give one when the second hardness maximum is reached. The next task is to link the aging rate with the strength and deformation properties of the material. This can be done through the hardness of the alloy, since indentation of the indenter is the same process of failure associated with large plastic strains [18].

In the absence of a load for triple alloys (2.5% Cu, 1.14% Mg and 3.0% Cu, 1.36% Mg), similar in composition to D16 T and AK4-1 T1 alloys of the same system, Hardy obtained a Q_p value of 32 and 33 kcal/mol, respectively [22]. This is approximately equal to the value of the activation energy of diffusion of copper in aluminum (32.6 kcal/mol [25], 1.4 ± 0.1 eV or 31.8 ± 2.3 kcal/mol [26]). After processing the reference data on changes in the deformation characteristics of AK4-1 T1 alloy specimens under monotonous loading after different aging modes, we obtained an estimate of $Q_p = 32.4$ kcal/mol and $\tau_{p0} = 10^{-10}$ s. That is, all estimates of the activation energy of the decay process of a solid solution turn out to be quite close.

After the introduction of the hardness function into the equation of the failure rate (5), with which the activation parameters of the fracture process are associated, and then calculations of the thermo-cyclic loading of structural elements were performed, taking into account the decay of a supersaturated metallic solid solution in these precipitation hardening alloys. Calculations of the fatigue failure process at low temperatures do not require taking into account such structural transformations, and it is quite acceptable to assume the structure of the material corresponding to the initial state of the alloy [18].

4. Examples of applying an interdisciplinary approach to practical tasks

Having the activation parameters U_0 and γ (which correspond to parameters A and B in Figure 5), it is possible to perform calculations for those loading conditions when the material flows throughout the entire volume, regardless of how the stresses and temperature change. The internal stresses in the so-called "fracture centers" naturally change, and this requires special modeling. Figure 10 shows the comparison of experimental data with the calculation for different temperature-time and temperature-force loading conditions of structural specimens made of AK-1 T1 alloy. Vertical lines correspond to the actual scatter of durability in the experiment, horizontal lines to the range of calculated estimates made taking into account the basic errors of the test program by load and temperature.

The figure shows that the calculated estimates of durability fall within a twofold range of deviations from the experimental data, which is usually observed when testing the same material of different batches. The calculations are made taking into account the decay of supersaturated metallic solid solution in a given alloy,

aged to the second maximum hardness (T1 state), representing the parameter $\dot{\epsilon}_0$ in (6) as the product of the residual strain by the frequency multiplier $\epsilon_* v_0$ and relating it to the change in the hardness of the alloy. In all cases, fracture occurs as a result of creep, regardless of how the stresses in material or the dangerous places of structural components change [4, 18].

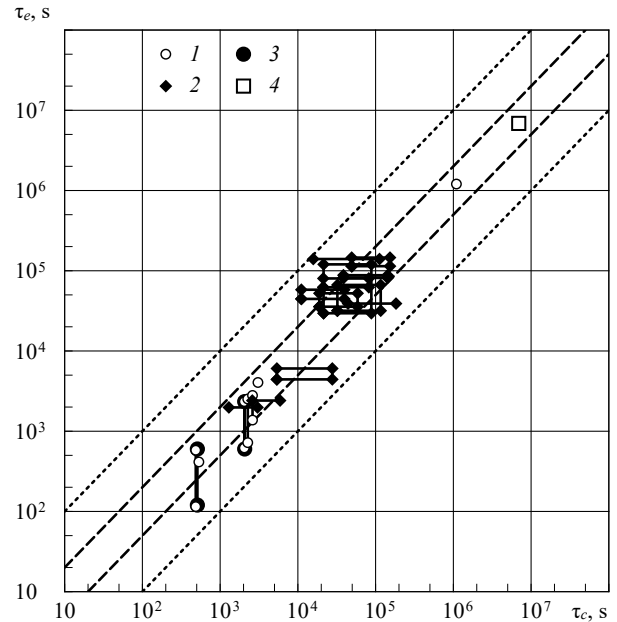


Figure 10: Comparison of calculated (τ_c) and experimental (τ_e) values of the durability of specimens and structural components made of AK4-1 T1 alloy tested at various loads and temperatures: 1, 2 – strip with a central hole and a longitudinal stringer under thermo-cyclic loading [4]; 3 – rod: constant and variable stresses at 543 K at 10 Hz [18]; 4 – full-scale structure fracture, tested under the specified temperature–force program (with addition of crack propagation period)

The determination of the remaining parameters of the structural model of the material (Fig. 5) requires cyclic loading at a constant mean stress component of the cycle σ_m . The values of temperature, frequency and shape of the loading cycle must be set. Stepwise increasing the amplitude of loading, we obtain the amplitude dependence of inelasticity (Fig. 8), which is used to select amplitude values for fatigue tests according to characteristic points. That is, for example, for the AB and BC ranges, two amplitude values must be selected each. Then, these modes must be tested with two mean load components. For each amplitude range, it is sufficient to know for any one mode the inelastic strain in the loading cycle. After parametric identification of the model, it is possible to perform calculations at a different temperature, frequency, cycle shape and generally at arbitrary changes in them, if one assumes that no changes in the structure occur in the material. Otherwise, this requires a separate study, and the material model parameters must be replaced by functions that represent these changes.

Using the relationship between inelastic strains and damage accumulation, the mathematical model makes it possible to calculate the durability for various spectra of external effects, be it stress or temperature, representing their implementation by piecewise linear approximation. Having solutions of differential equations, for example, (9) and for other structural elements of the

model at constant stresses or strains and linearly varying, for example, (15) and (16), it is possible to calculate any arbitrary process of temperature-force loading [4].

In Figure 11 shows a comparison of the calculated estimates of durability of structural components with experimental data for various loading cases. All tests were carried out at a temperature of 293 ± 2 K. Calculations were performed for a temperature of 293 K, assuming the structure of the material unchanged, corresponding to its initial state.

As in the previous example (Fig. 10), the calculated estimates of durability were made using a model of a design element that transforms in time the nominal stresses or loads into strains in the places of their concentration [4, 9]. For the specimens whose durability is marked by points 3, 4 and 6, the calculations were performed for two different quality batches of this material.

As in the previous example (Fig. 10), the calculated estimates of durability were made using a model of a design element that transforms in time the nominal stresses or loads into strains in the places of their concentration [4, 9]. For the specimens whose durability is marked by points 3, 4 and 6, the calculations were performed for two different quality batches of this material.

The time step of calculations at wide-band load spectrum is chosen minimum 0.25 or 0.5 period of the highest-frequency component of the spectrum. All load spectra were represented by equivalent polyharmonic pseudo-random processes (PRP) having the same spectral density, or by a real loading process recorded in operation [27]. The degree of discreteness of the spectrum depends on the material and type of the design element.

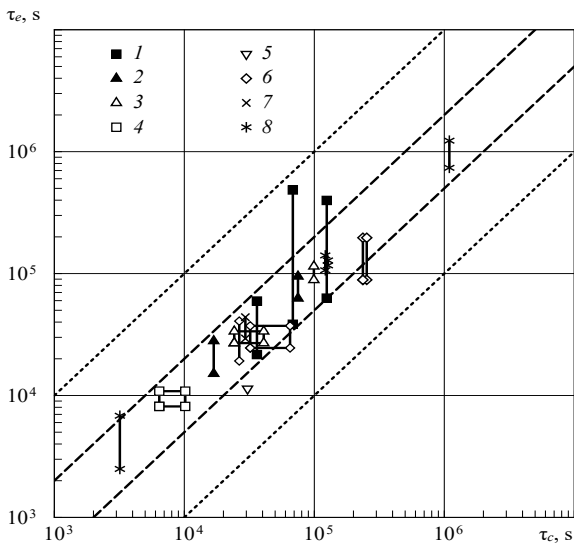


Figure 11: Comparison of calculated (τ_c) and experimental (τ_e) values of durability of structural specimens and structures made of 1201 T1 alloy tested under different loading programs ($T = 293 \pm 2$ K): 1 – plate-bar without notch, constant spectral density value in the interval $0.5 \div 10.5$ Hz, 6 harmonics; 2 – plate-bar without notch, narrowband random noise in the interval $0 \div 5.5$ Hz, 13 harmonics; 3 – plate-bar with notch, narrowband random noise in the interval $0 \div 5.5$ Hz, 13 harmonics; 4 – plate-bar with notch, block 87-step program, triangular cycle shape at 10 Hz; 5 – acoustic tests of panels in the interval $0 \div 200$ Hz; 6 – notched plate-bar, forced flight cycle 1200 s [23], compiled from records of bending moments on the wing of an airplane-laboratory; 7 – notched plate-bar, forced flight cycle GAG at 0.025 Hz; 8 – 30 mm wide plate-bar with a central hole 20 mm, cyclic tests in the frequency range $0.1 \div 40$ Hz with different cycle shape of loading

As in the case of variable temperatures (Fig. 10), the calculated estimates of durability are located in the range of twofold deviations from their experimental values. The calculations were performed based on the average statistical data of the durability of one of the semi-finished products of this material. To do this, two values of the mean cycle stresses are taken for each amplitude which selected by the inelastic characteristics of the material.

In each case, the structures are loaded in a different way; therefore, obtaining estimates requires statistical data on the typical operation conditions. Calculations are performed with averaged statistical data on loading, i.e., the averaged spectral density of the processes, which is then transformed to a discrete spectrum by the method of summation of elementary random functions [18]. As a result, one obtains a PRP, which is statistically equivalent to a real random process.

Examples of calculations for various PRPs, compared with the experiment, are given in the article [5]. The same real loading spectrum was modeled by a different number of harmonics distributed in several ways by frequency. This shows the significant effect of changes in the dispersion of the process in the high-frequency part of the spectrum on the durability of structural components.

To distinguish creep from fatigue, the units of measurement of durability must be uniform. Any unit of measurement always has a physical justification and a reference value [28]. The unit of measurement "cycle" does not exist in any system of units of measurement and cannot exist, since in each case it has a different content. Therefore, it is possible to distinguish cyclic creep from fatigue only if the durability is expressed in units of time, that is, the way the failure process actually occurs. In Figure 12 shows the dependences of durability on tensile stresses at their constant value and at cyclic tension with different frequencies and constant $\sigma_{\min} = 40$ MPa. The abscissa shows the equivalent stresses σ_{eq} corresponding to the constants at which the durability has the same value in accordance with expression similar to (7), –

$$\int_0^{\tau} \dot{\omega}(t, \sigma) dt = \tau v_0 \exp\left(-\frac{U_0 - \gamma \sigma_{eq}}{RT}\right).$$

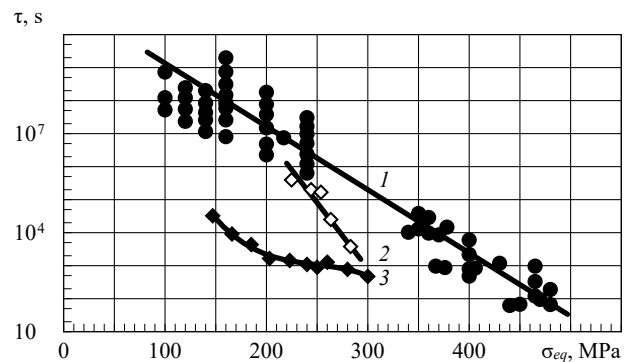


Figure 12: Durability of smooth specimens of AK4-1 T1 alloy tested at constant (1) and cyclically varying with a frequency of 0.05 (2) and 30 Hz (3) stresses (temperature 423 K)

The figure shows that at constant stresses, the logarithm of the durability linearly depends on the stresses, thereby illustrating the

main regularity of fracture (line 1). With alternating stresses varying with a low frequency (straight line 2), a decrease in the stress swing brings the value of the cyclic durability closer to the static one. It is clear that here we are dealing with cyclic creep, in which a decrease in durability occurs as a result of a concomitant relaxation of internal stresses, which decrease with a decrease in the loading rate [4, 6]. If the frequency is high (curve 3), a decrease in the stress swing increases the discrepancy between the durability under static and cyclic loading, which first increases and then becomes smaller, approaching the static durability at $\sigma_a \rightarrow 0$. With an increase in the stress swing, fatigue failure will be replaced by fracture from cyclic creep, and curve 3 intersects with straight line 2.

If the use of expression (5) is quite justified for a frequency of 0.05 Hz, then for fatigue failure at a frequency of 30 Hz, the approximating curve should be considered conditional (the lines in both cases are drawn according to the average logarithmic values of the durability). The process of failure during fatigue occurs in local volumes, the stresses in which are not known. In polymers, they can be evaluated by indirect methods, for example, by infrared spectroscopy [29]. In metal alloys and composite materials-structures, this can be done by inelasticity using mathematical models based on thermodynamic laws of fracture [4, 5, and 18].

Studies have targeted an interdisciplinary approach to the problem of the fracture of solids, for example, [30]. They consider in detail various aspects of the processes, depending on certain loading conditions, the materials used and their structures, and the mechanisms of the processes. A number of areas of knowledge related to the fracture of materials are considered and which explain what happens in this case. This is the knowledge necessary to understand the essence of the events taking place. Our approach does not consider the numerous details of the phenomena observed during fracture. This is like a cross-section of the whole problem in a certain plane. It is based on the fundamentals of the thermodynamics, and all types of fracture are considered from the same positions. The formulas given above are used specifically for calculations. Although for particular cases the mechanics hikes give quite satisfactory results, but fundamentally this does not solve the problem. The progress in solving complex strength problems is possible by combining the knowledge and methods of mechanics, physics and physical materials science.

The above and in article [1] examples show how large a body of information is provided by analysis of the rheological properties of materials and how many possibilities arise in modeling these properties if the mechanical models are filled with physical content. The approach outlined here, unlike the others currently in use, relies on a single conceptual framework: the notions of what a solid is and why it failures. Any material is an atomic-molecular system in thermal motion, and fracture occurs as a result of anharmonicity and stochasticity of the process of thermal vibrations of atoms in a solid body [7]. This is confirmed by a numerical experiment performed by the molecular dynamics method [31]. It follows that fracture is due to the internal energy of the solid, the measure of which is temperature [5, 7]. External effects only change the failure rate of a solid if they change its internal energy (thermal energy, electromagnetic radiation, chemical reactions). The end of the process is the achievement of a certain concentration of damage (microcracks, pores, delamination in composites), leading to its

transition to the next dimensional level [4, 5, 7, 9, and 14]. And the prediction of further fracture, such as crack propagation, is reduced to modeling the process of material failure at its tip by the same methods [7, 18, and 32].

5. Conclusion

The proposed methodology for predicting the durability of materials in structures shows that an interdisciplinary approach and reproduction in mathematical models of the processes of their failure and deformation as thermodynamic, allows us to solve those problems that have not yet been solved by methods of fracture mechanics. It is shown that it is based primarily on the study of the physical laws of fracture, which are revealed in experiments to determine durability at constant stresses and temperatures. Then these laws should be applied to other loading cases: variable loads and temperatures. And by involving the knowledge of materials science, it becomes possible to take into account in the calculation procedures other thermally activated processes accompanying the flow and failure of a solid, which changes its structure.

The above approach shows the practical sequence of actions required when investigating the strength properties of any new material. In order to study the strength, it is necessary, as a minimum, to test its specimens with different loading rates and at different temperatures, and then perform thermally activation analysis of the obtained data. To study the resource characteristics of a material, it is necessary to investigate their inelastic properties, comparing the results obtained with the endurance data, while taking into account the temporary nature of fatigue failure. This is followed by mathematical modeling of processes for solving problems of arbitrary temperature-force loading.

Conflict of Interest

The author did not have a conflict of interest with his colleagues using other approaches to the problem under consideration. The author experienced only a benevolent attitude, support and understanding.

Acknowledgment

The author is deeply grateful to H. D. Gringauz and N. A. Moshkin for a number of joint works that gave a lot of new, useful information and initiated the described approach, and with whose permission their results are presented here. He is also grateful to many other colleagues who, to one degree or another, helped him in numerous and laborious experiments.

References

- [1] M.G. Petrov, "Interdisciplinary approach to solving problems on the flow and fracture of materials," in XX International Conference on Methods of Aerophysical Research (ICMAR 2020), AIP Conference Proceedings 2351, doi: 10.1063/1.51004278.
- [2] W. Kauzmann, "Flow of solid metals from the standpoint of the chemical-rate theory," Transactions of the AIME, **143**, 57–83, 1941.
- [3] S. Glasstone, K.J. Laidler, H. Eyring, The theory of rate processes, McGraw-Hill, 1941.
- [4] M.G. Petrov, "Mathematical modeling of failure and deformation processes in metal alloys and composites," American Journal of Physics and Applications, **8** (4), 46–55, 2020, doi: 10.11648/j.ajpa.20200804.11.

- [5] M.G. Petrov, "Investigation of the longevity of materials on the basis of the kinetic concept of fracture," *Journal of Applied Mechanics and Technical Physics*, **62** (1), 145–156, 2021, doi: 10.1134/S0021894421010181.
- [6] V.A. Stepanov, N.N. Peschanskaya, V.V. Shpeizman, G.A. Nikonov, "Longevity of solids at complex loading," *International Journal of Fracture*, **11**, 851–867, 1975.
- [7] V.A. Petrov, A.Ya. Bashkarev, V.I. Vettegren, *Fizicheskiye osnovy prognozirovaniya dolgovechnosti konstruktivnykh materialov*, Polytechnika, 1993.
- [8] M.G. Petrov, A.I. Ravikovich, "Deformation and failure of aluminum alloys from the standpoint of kinetic concept of strength," *Journal of Applied Mechanics and Technical Physics*, **45** (1), 124–132, 2004.
- [9] M.G. Petrov, "Fundamental studies of strength physics – methodology of longevity prediction of materials under arbitrary thermally and forced effects," *International Journal of Environmental and Science Education*, **11** (17), 10211–10227, 2016.
- [10] M.G. Petrov, "Rol' protsessa polzuchesti svyazuyushchego v kinetike razrusheniya stekloplastika," in XII Mezhdunarodnoy nauchnoy shkoly-konferentsii: Fundamental'noye i Prikladnoye Materialovedeniye, Izdatelstvo AltGTU, 50–66, 2015.
- [11] J. Bailey, "An attempt to correlate some tensile strength measurements on glass," *Glass Industry*, **20**, 21–25, 1939.
- [12] M.G. Petrov, "Rheological properties of materials from the point of view of physical kinetics," *Journal of Applied Mechanics and Technical Physics*, **39** (1), 104–112, 1998.
- [13] A.R. Michetti, "Fatigue analysis of structural components through math-model simulation," *Experimental Mechanics*, **2**, 69–76, 1977.
- [14] S.N. Zhurkov, "Dilatonnyy mekhanizm prochnosti tvordykh tel," *Fizika Tverdogo Tela*, **25** (11), 33198–3323, 1983.
- [15] H.B. Dwight, *Tables of integrals and other mathematical data*, Macmillan Company, 1961.
- [16] A.J. Kennedy, *Processes creep and fatigue in metals*, Oliver and Boyd, 1962.
- [17] T-S. Kê, "Experimental evidence of the behavior of grain boundaries in metals," *Physical Review*, **71** (8), 533–546, 1947.
- [18] M.G. Petrov, *Prochnost' i dolgovechnost' elementov konstruktivnykh: podkhod na osnove modeley materiala kak fizicheskoy sredy*, Lambert Academic Publishing, 2015.
- [19] A.S. Nowick, B.S. Berry, *Anelastic relaxation in crystalline solids*, Academic Press, 1972.
- [20] M.L.V. Gayler, P. Parkhouse, "The ageing of high-purity 4 percent copper-aluminium alloy," *Journal of Institute of Metals*, **66**, 67–84, 1940.
- [21] M.L.V. Gayler, "The cold working of a high-purity aluminium alloy containing 4% of copper and its relation to age-hardening," *Journal of Institute of Metals*, **72**, 543–563, 1946.
- [22] H.K. Hardy, "The ageing characteristics of some ternary aluminium-copper-magnesium alloys with copper: magnesium weight ratios of 7 : 1 and 2.2 : 1," *Journal of Institute of Metals*, **83**, 17–33, 1954.
- [23] R. Graf, A. Guinier, "Influence de l'ecrouissage apres trempe sur les phenomenes de precipitation dans l'alliage aluminim-cuivre a 4% de cuivre," *Comptes rendus hebdomadaires des seances de l'academie des sciences*, **238**, 819–821, 1954.
- [24] J.W. Christian, *The theory of transformations in metals and alloys. Part I. Equilibrium and general kinetic theory*. 2nd ed., Pergamon Press, 1975.
- [25] L.H. Van Vlack, *Materials science for engineers*, Addison-Wesley, 1970.
- [26] A. Kelly, R.B. Nicholson, "Precipitation hardening," *Progress of Material Science*, Pergamon Press, **10**, 151–391, 1963.
- [27] M.G. Petrov, "On test programs of aircraft structures," in XVI International Conference on the Methods of Aerophysical Research (ICMAR 2012): abstracts, Part I, Kasan Federal University, 2012.
- [28] D. Kamke, K. Krämer, *Physikalische Grundlagen der Maßeinheiten*, B. G. Teubner, 1977.
- [29] V.I. Vettegren, I.I. Novak, K.J. Friedland, "Overstressed interatomic bonds in stressed polymers," *International Journal of Fracture*, **11**, 789–801, 1975.
- [30] T. Yokobori, *An interdisciplinary approach to fracture and strength of solids*, Wolters-Noordhoff Scientific Publications Ltd, 1968.
- [31] V.S. Yuschenko, E.D. Schukin, "Molekulyarno-dinamicheskoye modelirovaniye pri issledovanii mekhanicheskikh svoystv," *Fiziko-khimicheskaya mekhanika materialov*, **4**, 46–59, 1981.
- [32] V.R. Regel, A.M. Leksovskii, S.N. Sakiev, "The kinetics of the thermofluctuation – Induced micro- and macrocrack growth in plastic metals," *International Journal of Fracture*, **11**, 841–850, 1975.

Interpretable Rules Using Inductive Logic Programming Explaining Machine Learning Models: Case Study of Subclinical Mastitis Detection for Dairy Cows

Haruka Motohashi^{*1}, Hayato Ohwada²

¹Graduate School of Science and Technology, Department of Industrial Administration, Tokyo University of Science, Noda, Chiba, 278-8510, Japan

²Faculty of Science and Technology, Department of Industrial Administration, Tokyo University of Science, Noda, Chiba, 278-8510, Japan

ARTICLE INFO

Article history:

Received: 01 February, 2022

Accepted: 07 April, 2022

Online: 12 April, 2022

Keywords:

Inductive Logic Programming

Model Interpreting

Mastitis Detection

ABSTRACT

With the development of Internet of Things technology and the widespread use of smart devices, artificial intelligence is now being applied as a decision-making tool in a variety of fields. To make machine learning models, including deep neural network models, more interpretable, various techniques have been proposed. In this paper, a method for explaining the outputs of machine learning models using inductive logic programming is described. For an evaluation of this method, diagnostic models of bovine mastitis were trained using a dataset of dairy cows, and interpretable rules were obtained to explain the trained models. As a result, the rules obtained indicate that the trained classifiers detected mastitis cases depending on certain variations in the electrical conductivity (EC) values, and in some of these cases, the EC and lactate dehydrogenase fluctuated in different ways. The interpretable rules help people understand the outputs of machine learning models and encourage a practical introduction of the models as decision-making tools.

1 Introduction

This paper is an extension of a study originally presented at the 2020 IEEE 19th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC) [1].

With the development of Internet of Things technology and the widespread use of smart devices, artificial intelligence is now being used as a decision-making tool in a variety of fields. Moreover, various machine learning models have been proposed to support an efficient and accurate medical diagnosis. Such models are expected to not only detect patients correctly, but also reveal the basis of the diagnosis.

Various techniques have been proposed to make machine learning models, including deep neural network models, more interpretable. Decision-tree-based algorithms (e.g., random forest and lightGBM [2]) provide the feature importance based on the frequency of all features in the trees generated by the algorithms. Some algorithms approximate original complex models (including a deep neural network) locally with simpler interpretable models [3, 4]. For convolutional neural network used to solving image processing tasks, gradient-based highlighting represents important

regions in images where the networks focus to detect target objects or track them [5, 6].

Another approach to interpreting machine learning models is to describe their outputs using interpretable rules. Inductive logic programming (ILP) is based on predicate logic and can produce rules using inductive learning. ILP has the advantage of obtaining interpretable classification rules from training data and representing the opinions of domain experts [7, 8].

The interpretability of machine learning models has encouraged their introduction in decision-making applied in fields such as medical, including veterinary, diagnosis. Bovine mastitis, which is an inflammation of the udder or mammary gland owing to physical trauma or infection, is a common disease in dairy cattle, which dairy farmers must control to prevent economic losses.

With the introduction of auto milking systems, it has become easier to measure the indicators needed for cow health management during milking and to detect common diseases in dairy cows, including mastitis. Auto milking systems enable farmers to utilize auto mastitis detection using indicators such as milk yield, electrical conductivity, fat, protein, lactose and blood in the milk, and milk flow rate. In addition, SCC and various systems using statistical

*Corresponding Author: Haruka Motohashi, Faculty of Science and Technology, Tokyo University of Science, 2641 Yamazaki, Noda-shi, Chiba Prefecture 278-8510, Japan, Tel: +81-4-7124-1501, & 7420701@ed.tus.ac.jp

models and machine learning, including artificial neural networks, have also been reported [9]–[10].

In this study, we propose a method for explaining the outputs from machine learning models using ILP. For an evaluation of the method, diagnosis models of bovine mastitis were trained using a dataset of dairy cows, and interpretable rules were generated using ILP and the explained outputs of the mastitis detection model.

2 Method Used to Explain Classifiers through Interpretable Rules

An overview of the method used in this study is presented in Figure 1. This method aims to generate logic rules using background knowledge and outputs of a classifier, and to interpret the classification model. Interpretable rules are generated using ILP. In contrast to ordinary machine learning models such as deep neural networks, a resultant set of rules produced using ILP generally represents patterns in the given datasets. In this study, ILP is applied to outputs from classifiers trained using machine learning methods, and the set of rules generated describe how the model classifies the instances.

Although other methods based on a linear local approximation [3, 4] represent the importance of each feature, an explanation of machine learning models using ILP describes the models using nonlinear relationships with multiple features. Thus, methods for explaining classifiers can be applied to models with a complex architecture, such as deep neural networks [11, 12].

In this study, although the architectures of the machine learning models are maintained, the outputs of the models are given to an ILP system. Therefore, the proposed method is available regardless of the machine learning methods used for model training. Moreover, the definition of predicates used in ILP can be distinguished from the features in the classifiers and the predicates take advantage as a way to reflect knowledge of domain experts.

As shown in Figure 1, in the first step, a classification model is trained using machine learning methods and outputs of the classification are obtained. Background knowledge of the dataset and its outputs are then added into an ILP system, called Parallel GKS [13]. Finally, the classifier is interpreted based on the set of rules.

3 Case Study: Explanation of Bovine Mastitis Detection Model

To evaluate the proposed method, a classification model for the subclinical mastitis detection of dairy cows was trained using machine learning. With the introduction of auto milking systems, it has become easier to measure the indicators needed for cow health management during milking and to detect common diseases in dairy cows including mastitis.

Previous studies [9]–[10] used records of veterinary treatments and somatic cell count (SCC) for labeling the data of every milking as clinical or subclinical mastitis, and their models predicted the status of the quarters during each milking. SCC is generally used for the diagnosis of subclinical mastitis, and the most frequently used threshold for defining subclinical mastitis is 200,000 cells/mL [14].

However, it is thought that SCC can be affected by other factors such as the lactation number, stress, season, and breed [15].

As novel mastitis detection approaches, some biomarkers for mastitis detection have been discovered [16]–[17]. In particular, lactate dehydrogenase (LDH), which is related to inflammation, and according to a previous study is the biomarker with the lowest validation [18], is measured practically using a commercial milking machine. The result is applied to calculate the risk of developing mastitis as part of a milk analysis. However, mastitis detection using such biomarkers is still not a common approach for farmers because the equipment required is quite expensive. Therefore, in this study, a dataset in which cows labeled as either healthy or having subclinical mastitis based on the LDH values was prepared, and a common measurement, i.e., the electrical conductivity (EC), was used as a feature through the application of machine learning.

The dataset used in this study was collected between September 2018 and December 2021 at a farm in Hokkaido, Japan. On this farm, cows are milked any time they want, and items other than mastitis risk are measured during every milking. Data from September 2018 to August 2020 collected on the farm were used to train the detection model, and the remaining data were used to evaluate the model.

Datasets measured using an auto milking machine (a DeLaval Voluntary Milking System™; VMS) and a milk analyzer (a DeLaval Herd Navigator™; HN) were used. The HN measures the LDH, which is an index of subclinical mastitis, in milk and is used to calculate the risk of contracting mastitis.

The mastitis risk takes a value of zero to 100. On a farm, an HN measurement of greater than 70 allows farmers to suspect that a cow has mastitis. Therefore, in this study, subclinical mastitis cases were determined based on the mastitis risk. If her mastitis risk is above 70, the cow has subclinical mastitis; otherwise, her udder is disease-free.

3.1 Data Preprocessing

In this detection model, two features are calculated based on the EC obtained from VMS. One is the maximum EC values (max_EC) in the udder, and the other is ratio of the maximum to minimum EC values, i.e., the inter-quarter ratio (IQR). The EC is one of the measurements related to mastitis, and its value increases when a cow has mastitis [19]. According to our previous study [1], this mastitis detection model includes a four-day time series of these two features from three days prior to the prediction day, with eight features in total.

As mentioned above, cows on this farm are milked any time they want, and items without mastitis risk are measured during every milking. In addition, LDH in milk (indicating the risk of mastitis) is generally measured once every day to every three days, and the next measurement day is determined depending on the risk value. Therefore, this dataset consists of time series data, and the number of data points of a cow differs each day. Thus, features are calculated using data on the milking with the highest EC.

The preprocessed datasets in this study are shown in Table 1. In the test dataset, labeled samples are used for an evaluation of the detection model, and unlabeled samples are only used for generating rules through ILP.

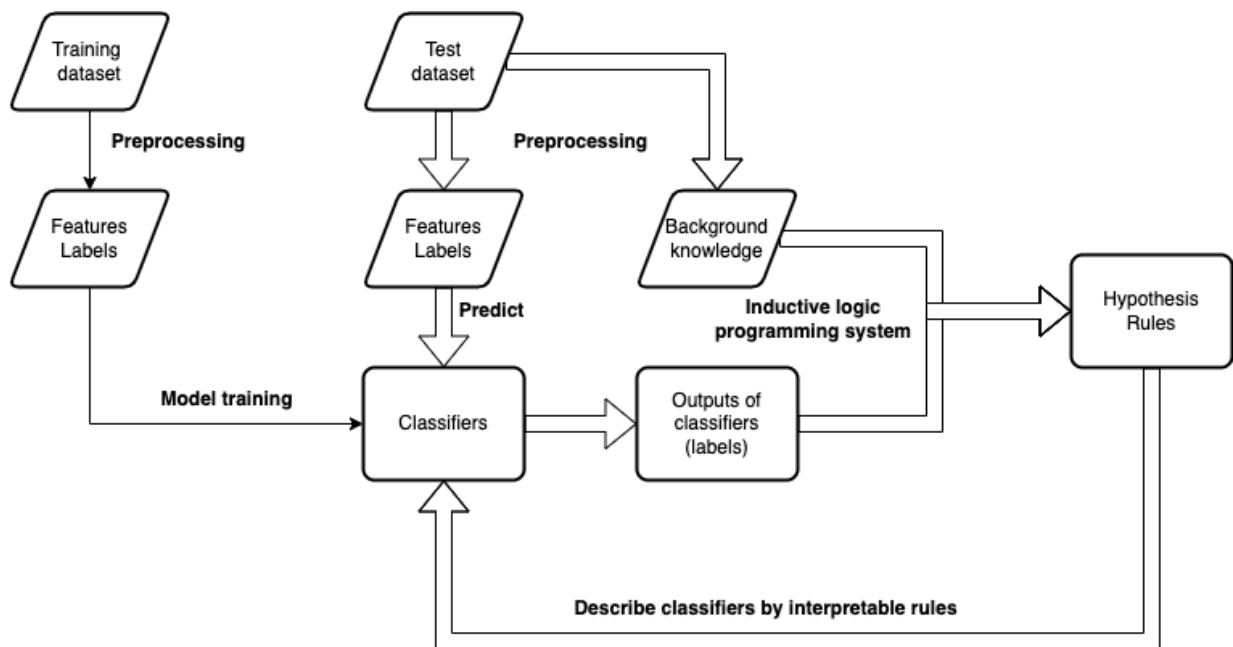


Figure 1: Overview of our method used to explain classifiers through interpretable rules.

Table 1: Number of cows and samples in the training and test datasets.

Dataset	Training	Testing
Data extraction period	2018-9-23 – 2020-11-1	2020-11-2 – 2021-12-2
Samples	98825	122372
Subclinical mastitis	4176	2121
Fine	94649	50771
unlabeled	-	69480

3.2 Learning Classification Models

After data preceding, classifiers for bovine mastitis detection are trained using a support vector machine (SVM). In this case, cows with subclinical mastitis are sparsely present in the dataset, and hence under-sampling (using the repeated edited nearest neighbors algorithm [20, 21]) was applied to the training dataset, and a regularization parameter in SVM (C) was adjusted using class weights, which are inversely proportional to the class frequencies in the input data.

The trained models were evaluated through a 10-fold cross validation using the sensitivity, specificity, and area under a receiver operating characteristic curve, as in previous studies on mastitis detection [9, 22, 23]. To evaluate the small number of false positives, the precision was also used for the evaluation.

Samples in the test datasets were given to the trained dataset and classified as fine or having mastitis. Using these outputs, interpretable rules explaining the detection model were generated using ILP.

3.3 Generation of Rules Using ILP to Explain the Trained Models

Using the outputs from the trained classifiers and background knowledge of samples in the test dataset, interpretable rules were obtained through ILP learning. Like other machine learning methods, ILP algorithms extract patterns in the samples with a certain label. In this study, an ILP system called GKS [24, 13] was used to employ ILP and generate rules.

Background knowledge is used in ILP learning, similar to features in other machine learning methods, and is represented by predicates in terms of the logic program. To describe the numerical features in terms of ILP, features were discretized based on the definition of the predicates for background knowledge and rules. In this case, three arguments were defined, as shown in Table 2. Such background knowledge of samples labeled by the trained classifier was given to the ILP system, and rules consisting of the predicates were generated.

A variable id in Table 2 represents one sample in the datasets. @IQR is a predicate which describes difference of the EC values between cows' quarters on the prediction day and whose variable, IQR, takes four values ($\leq mean - std$, $> mean - std$, $> mean$, $> mean + std$ where $mean$ is 1.08 and std is 0.07) defined by discretized values of IQR (features in the detection models).

Table 2: Definition of the predicates in the subclinical mastitis detection model.

Predicate	Definition
@IQR,+id,+IQR @IQR,+id,-IQR @IQR,+id,#IQR	the maximum value of EC / the minimum value of EC in quarters (discretized by three thresholds: <i>mean - std, mean, mean + std</i>)
@delta_maxEC,+id,+day1,+day2,#delta @delta_maxEC,+id,-day1,+day2,#delta @delta_maxEC,+id,#day1,+day2,#delta @delta_maxEC,+id,+day1,-day2,#delta @delta_maxEC,+id,-day1,-day2,#delta @delta_maxEC,+id,#day1,-day2,#delta @delta_maxEC,+id,+day1,#day2,#delta @delta_maxEC,+id,-day1,#day2,#delta @delta_maxEC,+id,#day1,#day2,#delta	the difference of max_EC between two days (<i>minus, flat, plus</i>)
@before_day,+day1,-day2	sequence of days (day_0 is the prediction day and day_n is the day following day_{n+1})

@delta_maxEC represents the difference of max_EC values between two consecutive days (from three days prior to the prediction day) and a variable delta takes three values (*minus, flat, plus* which describes the value of max_EC decreases or increases by over 0.4 or remain flat).

@delta_maxEC also has two variables representing targeted days from three days prior to the prediction day and these variables take four values ($day_3, day_2, day_1, day_0$). @before_day is a predicate which expresses an ordinal relation between these four values. This predicate contributes to generate rules flexibly, which consider difference of the values between arbitrary consecutive two days and mention changes of the values between variable periods before the prediction day, unlike tree-based algorithms.

Outputs from the learned classifier and the background knowledge were given to parallel GKS and ILP learning was employed. Finally, interpretable rules for mastitis detection models were obtained.

Table 3: Subclinical mastitis detection performance in the training dataset (a 10-fold cross validation was used) and the test dataset.

	Sensitivity	Specificity	Precision	AUC
Training	0.668	0.814	0.137	0.809
Test	0.667	0.840	0.148	0.831

4 Result and Discussion

To evaluate our method for explaining machine learning models using interpretable rules, it was applied to the classification problem of bovine mastitis detection. Table 3 lists the evaluation results for the classifier of subclinical mastitis detection using an SVM. In the test dataset, 79 records of veterinary treatment were included, and 66 out of 79 (83.5%) records were detected as subclinical mastitis by the classifier, whereas the target label of the model underestimated

the mastitis risk. Therefore, the trained model detected some of the clinical mastitis cases correctly, although the precision of the classifier was 15%.

Using outputs from the classifier, ILP learning was employed, and interpretable rules were obtained. The rules that were the most readable and related to the real conditions of mastitis are listed below. These rules describe the relationship between mastitis and the variation of the EC values.

- Rule1 pos(A) :- IQR(A, > mean + std), delta_maxEC(A, B, C, plus), delta_maxEC(A, day0, day1, flat)
- Rule2 pos(A) :- delta_maxEC(A, B, C, minus), delta_maxEC(A, C, D, minus)
- Rule3 pos(A) :- IQR(A, > mean), delta_maxEC(A, day2, B, plus)
- Rule4 pos(A) :- delta_maxEC(A, B, C, minus), delta_maxEC(A, day2, day3, plus)

These rules were described variation of max_EC values before rising mastitis risk by combination of @delta_maxEC. Values A, B and C given to variables day1 or day2 in this predicate represented arbitrary consecutive days before the prediction day, which made generated rules more scalable.

Among the rules obtained, some clearly explaining the classifier are described in detail.

- Rule1 pos(A) :- IQR(A, > mean + std), delta_maxEC(A, B, C, plus), delta_maxEC(A, day0, day1, flat)

Rule 1 indicates that cows whose IQR value is extremely high and whose max_EC values continuously increase have a high risk of subclinical mastitis. Figure 2 is a case corresponding to Rule 1. This rule also indicates that the detection model comprehends the typical relationship between bovine mastitis and electrical conductivity of milk. In this case, EC and LDH increased simultaneously, and the classifier detected a high mastitis risk as the mastitis risk alarm based on LDH.

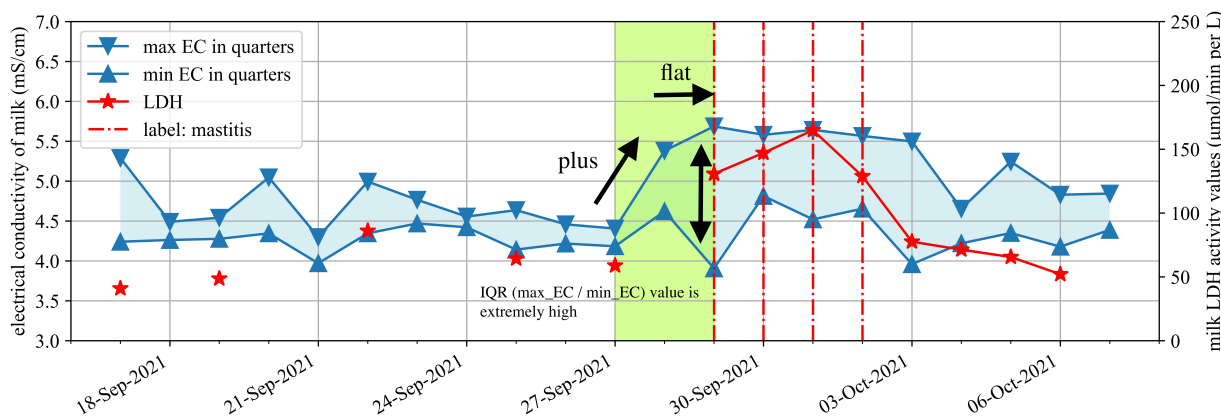


Figure 2: One of the subclinical mastitis cases corresponding to Rule 1.

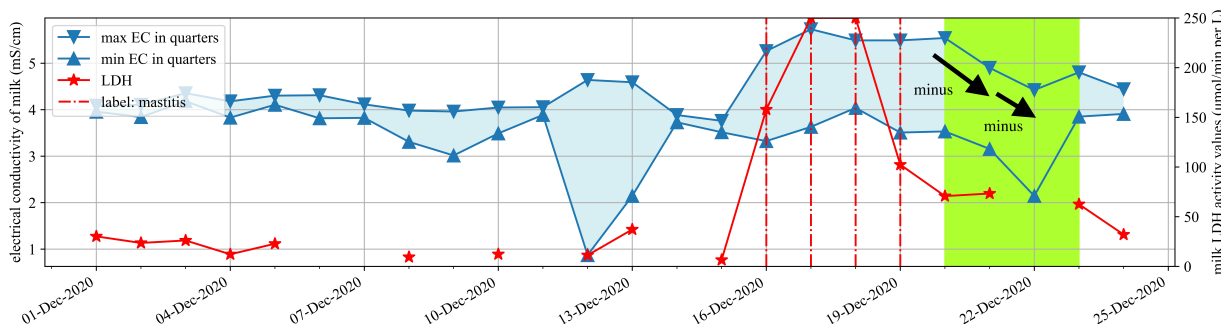


Figure 3: One of the subclinical mastitis cases in which the values of EC and LDH changed differently corresponding to Rule 2.

Rule2 pos(A) :- delta_maxEC(A, B, C, minus), delta_maxEC(A, C, D, minus)

Rule 2 indicates that cows whose max_EC values have been decreasing for two straight days have a risk of subclinical mastitis. Rule 2 apparently conflicts with Rule 1. However, this rule explains some cases of cows with subclinical mastitis. Figure 3 shows an example of cases in which the values of EC and LDH increased at disparate times. In this case, the mastitis risk values based on LDH, i.e., the target label of the classifier used in this study, become lower before the EC values began decreasing. However, the cow was deemed to be disordered by the farm staff and received veterinary treatment on December 23rd. Therefore, Rule 2 suggests that the trained classifier detected clinical cases that continued after the LDH values began decreasing.

The rules generated by ILP practically interpreted the mastitis detection model in this study and provided explanation of the mastitis detection, which were available for users of this detection system, staff of the farm, to understand how to classify cows as mastitis or fine. Providing reason of diagnosis by machine learning to users of the models would accelerate to develop the models as well as support farmers control health of dairy cows efficiently.

5 Conclusion

In this study, a method for explaining the outputs from machine learning models using ILP was suggested and evaluated when applied to the task of subclinical mastitis detection for dairy cows.

For an earlier detection of the onset of subclinical mastitis, a model for subclinical mastitis detection trained using risk values based on LDH was proposed. Interpretable rules were then generated using ILP to interpret the trained models. The rules obtained indicate that the trained classifiers detect mastitis cases depending on a certain variation of the EC values and that the EC and LDH fluctuate in different ways. The interpretable rules help in understanding the outputs of machine learning models and encourage a practical introduction of models as tools for decision making.

References

- [1] H. Motohashi, H. Ohwada, C. Kubota, "Early detection method for subclinical mastitis in auto milking systems using machine learning," in 2020 IEEE 19th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), 76–83, IEEE, 2020, doi:10.1109/iccicc50026.2020.9450258.
- [2] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, "Lightgbm: a highly efficient gradient boosting decision tree," in Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, 3149–3157, Curran Associates Inc., Red Hook, NY, USA, 2017, doi:10.5555/3294996.3295074.
- [3] M. T. Ribeiro, S. Singh, C. Guestrin, "Why should I trust you?" Explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 1135–1144, 2016, doi:10.1145/2939672.2939778.
- [4] S. M. Lundberg, S.-I. Lee, "A unified approach to interpreting model predictions," in Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, 4768–4777, Curran Associates Inc., Red Hook, NY, USA, 2017, doi:10.5555/3295222.3295230.

- [5] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, "Grad-cam: visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 618–626, 2017, doi:10.1109/iccv.2017.74.
- [6] D. Smilkov, N. Thorat, B. Kim, F. Viégas, M. Wattenberg, "Smoothgrad: removing noise by adding noise," arXiv preprint arXiv:1706.03825, 2017, doi:10.48550/arXiv.1706.03825.
- [7] N. P. Martono, K. Abe, T. Yamaguchi, H. Ohwada, "An analysis of motion transition in subtle errors using inductive logic programming: a case study in approaches to mild cognitive impairment," *International Journal of Software Science and Computational Intelligence (IJSSCI)*, **10**(1), 27–37, 2018, doi:10.4018/ijssci.2018010103.
- [8] S. Sasaki, R. Hatano, H. Ohwada, H. Nishiyama, "Estimating productivity of dairy cows by inductive logic programming," in *Proceedings of The 29th International Conference on Inductive Logic Programming*, 2019, doi:10.1007/978-3-030-49210-6.
- [9] D. B. Jensen, H. Hogeveen, A. De Vries, "Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis," *Journal of Dairy Science*, **99**(9), 7344–7361, 2016, doi:10.3168/jds.2015-10060.
- [10] W. Steeneveld, L. C. van der Gaag, W. Ouweltjes, H. Mollenhorst, H. Hogeveen, "Discriminating between true-positive and false-positive clinical mastitis alerts from automatic milking systems," *Journal of Dairy Science*, **93**(6), 2559–2568, 2010, doi:10.3168/jds.2009-3020.
- [11] J. Rabold, M. Siebers, U. Schmid, "Explaining black-box classifiers with ILP—empowering LIME with Aleph to approximate non-linear decisions with relational rules," in *International Conference on Inductive Logic Programming*, 105–117, Springer, 2018, doi:10.1007/978-3-319-99960-9_7.
- [12] J. Rabold, G. Schwalbe, U. Schmid, "Expressive explanations of dnns by combining concept analysis with ilp," in *KI 2020: Advances in Artificial Intelligence*, 148–162, Springer International Publishing, Cham, 2020, doi:10.1007/978-3-030-58285-2_11.
- [13] H. Nishiyama, H. Ohwada, "Parallel inductive logic programming system for superlinear speedup," in *International Conference on Inductive Logic Programming*, 112–123, Springer, 2017, doi:10.1007/978-3-319-78090-0_8.
- [14] S. Pyörälä, "Indicators of inflammation in the diagnosis of mastitis," *Veterinary research*, **34**(5), 565–578, 2003, doi:10.1051/vetres:2003026.
- [15] A. J. Schepers, T. J. G. M. Lam, Y. H. Schukken, J. B. M. Wilmink, W. J. A. Hanekamp, "Estimation of variance components for somatic cell counts to determine thresholds for uninfected quarters," *Journal of Dairy Science*, **80**(8), 1833–1840, 1997, doi:10.3168/jds.s0022-0302(97)76118-6.
- [16] C. M. Duarte, P. P. Freitas, R. Bexiga, "Technological advances in bovine mastitis diagnosis: an overview," *Journal of Veterinary Diagnostic Investigation*, **27**(6), 665–672, 2015, doi:10.1177/1040638715603087.
- [17] Y. C. Lai, T. Fujikawa, T. Maemura, T. Ando, G. Kitahara, Y. Endo, O. Yamato, M. Koiwa, C. Kubota, N. Miura, "Inflammation-related microRNA expression level in the bovine milk is affected by mastitis," *PLoS One*, **12**(5), e0177182, 2017, doi:10.1371/journal.pone.0177182.
- [18] M. Åkerstedt, L. Forsbäck, T. Larsen, K. Svennersten-Sjaunja, "Natural variation in biomarkers indicating mastitis in healthy cows," *Journal of Dairy Research*, **78**(1), 88–96, 2011, doi:10.1017/S0022029910000786.
- [19] E. Norberg, H. Hogeveen, I. R. Korsgaard, N. C. Friggens, K. H. M. N. Sloth, P. Løvendahl, "Electrical conductivity of milk: ability to predict mastitis status," *Journal of Dairy Science*, **87**(4), 1099–1107, 2004, doi:10.3168/jds.s0022-0302(04)73256-7.
- [20] I. Tomek, "An experiment with the edited nearest-neighbor rule," *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-6**(6), 448–452, 1976, doi:10.1109/TSMC.1976.4309523.
- [21] D. L. Wilson, "Asymptotic properties of nearest neighbor rules using edited data," *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-2**(3), 408–421, 1972, doi:10.1109/TSMC.1972.4309137.
- [22] M. Khatun, P. C. Thomson, K. L. Kerrisk, N. A. Lyons, C. E. F. Clark, J. Molfino, S. C. García, "Development of a new clinical mastitis detection method for automatic milking systems," *Journal of Dairy Science*, **101**(10), 9385–9395, 2018, doi:10.3168/jds.2017-14310.
- [23] D. Cavero, K.-H. Tölle, C. Henze, C. Buxadé, J. Krieter, "Mastitis detection in dairy cows by application of neural networks," *Livestock Science*, **114**(2-3), 280–286, 2008, doi:10.1016/j.livsci.2007.05.012.
- [24] F. Mizoguchi, H. Ohwada, "Constrained relative least general generalization for inducing constraint logic programs," *New Generation Computing*, **13**(3), 335–368, 1995, doi:10.1007/bf03037230.

COVIDFREE App: The User-Enabling Contact Prevention Application: A Review

Edgard Musafiri Mimo^{1,*}, Troy McDaniel¹, Jeremie Biringanine Ruvunangiza²

¹ *Arizona State University, Systems Engineering, The Polytechnic School, Mesa, 85212, USA*

³ *University of Mons, Computer Science, The Polytechnic School, Mons, 7000, Belgium*

ARTICLE INFO

Article history:

Received: 11 February, 2022

Accepted: 16 March, 2022

Online: 19 April, 2022

Keywords:

COVIDFREE App

Location-based services

Social distancing

Covid-19

Location tracking

ABSTRACT

The use of Covid-19 contact tracing applications has become almost irrelevant now that several flavors of Covid-19 vaccine have been developed and are constantly being distributed to people during the pandemic to help alleviate the need for lockdowns. Also, the availability of at-home testing kits and testing sites means that people do not need to contact trace as much since individuals can get tested and follow the health guidelines in ensuring their health and the safety of others around them. Nevertheless, governments around the world are still faced with the Covid-19 pandemic challenge because the virus is not yet controlled due to the different variants and the rapid contamination rate that outpaced the logistic supply chain processes in the distribution of the vaccines and the time it takes in convincing individuals to take the vaccine swiftly to reach herd immunity. Therefore, the current pressing need is that of addressing the infection rate by finding ways and solutions to minimize or slow down contamination among people especially with the increased number of variants. This paper is an extension of the "COVIDFREE App: The User-Enabling Contact Prevention Application" work originally presented in 2020 IEEE International Symposium on Technology and Society (ISTAS) conference that provided a smartphone application architecture with the goal of proactively enabling users to avoid encountering infected Covid-19 patients. This paper elucidates and discusses additional concerns not thoroughly addressed previously regarding the Covid-19 variants, vaccines, booster, and infection rates, and demonstrates the feasibility of the proposed architecture with a web application prototype. This paper also discusses the benefits of funding and developing contact tracing and prevention applications, such as the COVIDFREE App, to provide the needed ingredient in reducing the infection rate and provide citizens the needed preparedness and relief in actively fighting the virus.

1. Introduction

The world has made so many financial, social, political, and technological efforts and investments in fighting and controlling the Covid-19 virus spread as a remedy to end the virus fueled pandemic the world was forced in. All the efforts were made as more information about the virus became available starting with how one can avoid catching the virus to what one must do once infected to recover and not spread the virus to others and so on and so forth. The declared Covid-19 International Public Health Emergency Issue that began in 2020 has drastically changed the world, and there is no going back to normal anymore rather there is a contemplation of adopting a new normal. Covid-19 is a highly contagious respiratory virus with several unique features that have

been identified with the ever-increasing number of infection cases to date.

With more detail about Covid-19 available today on how some infected individuals do not develop symptoms right away, while others stay asymptomatic throughout the duration of their treatment, and others are able to catch the different mutation and variation of the virus, it is important to evaluate all the alternative remediation solutions that can be instigated to manage the pandemic health crisis more efficiently. Several governments around the world have responded by exploring and seeking to harness technology in the fight against this fatal illness. Covid-19 has influenced the lives of almost everyone on the earth by forcing governments around the world to implement lockdowns and sanctions, and in certain situations, enforce measures that involve

* Corresponding Author: Edgard Musafiri Mimo, emusafir@asu.edu

work-from-home regulations, implement strong physical distancing safeguards, and set up emergency health responses that necessitate thorough rearrangement to perform mass testing, patient managing, and care giving [1].

These efforts are aimed at containing and controlling the virus's transmission until definitive cures or vaccines are produced and widely administered. Considering the effort and investment in producing the vaccines and the vaccine boosters, and getting people vaccinated and boosted, there is still a lot of work to be done in combating the virus and ensuring people's safety [2]. As a result, it is evident that the Covid-19 testing and vaccine solutions are maybe the remedy for the virus, but not the remedy of its infection rate. They simply cannot keep up with the pace of transmissibility of the virus as the supply chain logistic processes are also impacted by the pandemic. Hence, there is a need to address the contamination issues among the people by ensuring the prevention of contacts or encounters among people proactively because potential Covid-19 variants may continue to develop due to the constant virus reproduction among humans [3].

This is the intention of the COVIDFREE app because governments all over the world are inclined to adopt and employ mobile contact tracing and awareness applications to automatically manage, trace, and investigate recent interactions of both the newly tested Covid-19 infected individuals and those that are recovering from it regardless of vaccine status [4]. The potential use of such mobile and web applications has produced numerous discussions surrounding privacy, security, data management, contact projection algorithms, and cyber-attack vulnerabilities. The previously proposed smartphone application architecture that COVIDFREE employs aims to minimize the above-mentioned concerns by using only minimal information about the users and provide abstraction layers to ensure the security and the privacy of users are preserved in the process of giving the users more control and privacy of their data [1].

The COVIDFREE application aims to enhance users' situational mindfulness through communication of unsafe locations and proactively prompt them to avoid these locations while dynamically taking into the account the daily reported infection rate per location. Additionally, the COVIDFREE application provides the means of customizing by allowing users the control of considering their Overall Risk Density Safety Factor in the risk calculation based on health requirements and numerous customizable user-specific situations [1]. This paper discusses a prototype of a previously proposed proactive smartphone app with a centralized approach that aids uninfected people regardless of their vaccination status to overcome the stress and fear of getting contaminated by improving their situational awareness. The proposed proactive smartphone application enables users to informatively avoid congested places and tested infected individuals because the application notifies them whenever they are within 10 to 50 feet (depending on parameters) of an anonymous, confirmed infected person [1].

2. Insights and Concerns

2.1. General Anxieties

There are several unsolved questions about how technology can help provide a remedy in addressing the Covid-19

transmissibility issues even now that the vaccines and vaccine boosters are available, since they are not yet distributed everywhere due to their demands and the citizens' decision in taking them. The questions surrounding the quarantine period and the habits and behaviors of citizens in how they choose to live their lives, whether it is by masking up or just limiting their movements outside of their safe locations, cannot be overlooked. The simple fact that many Covid-19 patients are asymptomatic leaves the world with so little options to consider in providing the remediation needed to enable the minimization of the transmission of the virus from one patient to others [1], [5].

If some patients do not have symptoms right away and some are symptoms free while carrying the virus, then the Covid-19 infection rate would remain a burden to carry as more undercover carriers could prove difficult to identify and avoid [5]. As a result, Covid-19 testing for asymptomatic individual is problematic since infected asymptomatic individual may have already spread the virus to numerous others before being tested. Consequently, it is critical to guarantee that individuals feel secure going to and returning from testing centers by understanding how infectious their surrounding is. For instance, regardless of symptoms, it is beneficial for users to know how close they may be to someone who is contaminated and has had their infection confirmed by a previously taken Covid-19 test.

2.2. Tracking Complexity

The complexity of Covid-19 spread is vast, so there is a need of a proactive response strategy that locates confirmed infected people outside their safe location and provides means of avoiding them anonymously without any form of profiling through direct notification of users' surrounding exposure range [1], [6]. This proactive strategy is urgently needed to ensure people feel safer and more educated about their environments when it comes to Covid-19 health risks. We believe that proactively notifying people to avoid encountering a person that has tested positive for Covid-19 will help minimize the infection rate and allow the vaccines and boosters' effects to be effectively noticed as herd immunity is being achieved.

Reactive techniques of contact tracing or a passive approach of waiting on the vaccines and boosters only, would not provide the rapid remedy that people are waiting for as long as the contamination rate stays high and unaddressed in a proactive manner that of relying on informative individual discretionary social distancing and isolation. The only possibilities left imminent involve the tradeoff between what privileges one can live with and what necessities one cannot live without as new Covid-19 demands need to be satisfied and new habits developed to cope with the world's present changes of the new normal. Hence, it is necessary to provide tools that give users insights and proactive prompt notifications to avoid the symptoms free carriers and minimize the potential risk of catching the virus. This is a great way to make use of the collected testing daily data to ensure the location of the individuals that test positive are activated until they test negative depending on their vaccination status and their quarantine periods. It will ensure other users would promptly avoid them should interactions' occasions arise.

Therefore, the method of preventing anyone from getting the virus from the afflicted person regardless of their vaccine status

that the COVIDFREE application provides, ensures people are informed and at peace with their movements while minimizing their concerns regarding the virus and their health and safety. With a couple of years in the pandemic, there are still some concerns regarding Covid-19 transmission and propagation that are yet to be completely understood regarding the variety of ways the virus can spread [6]. Thus, it is critical to access the user's physical surroundings, social connections, and health to determine the likelihood of being contaminated. The virus's complicated transmission characteristic generates a lot of concerns among the people. As a result, recommendations to maintain social distance and wear personal protective equipment (PPE) help to limit the virus's transmission. The case that someone can get the virus on their way to the testing site and then obtain a negative test report to become an undercover carrier in the testing process. Being uninformed about one's immediate surrounding adds to the concerns and does not help answer the question of when the virus is spread especially with the case of vaccinated individuals' potential to be contaminated again.

Another issue that makes the concerns worse is the amount of time the virus takes to disperse in the air, which varies based on area and spaces' ventilation systems. Thus, without the full visibility of the virus' transmitters, it's hard to identify, track, and locate unsafe sites accurately without depending on the tracking of individuals who have tested positive. This is a way that can help provide insights to address the question of where someone can potentially get Covid-19 by being in the vicinity of an active carrier (person to person) rather than a passive one (via the environment). It is convoluted to make the right informed decision regarding the virus's transmissibility during travel if critical information and notifications are not available to people in real time to support informed travel experiences.

3. Simplified Prototype Architecture

The overall density safety factor proposed earlier must be adjusted accordingly when considering the vaccine safety factor to determine how to account for the efficacy of the vaccine and their booster to provide a realistic vaccination safety factor [1]. The proposal may be using the normalized reported average efficacy value per age group from a credible vaccine approval institution like the Food and Drug Administration (FDA) agency. Nevertheless, the previously proposed normalized overall density safety factor along with all the considered factors as discussed in this paper remain a great indicator and tool to dynamically adjust per users' needs the notification distance range to facilitate users in avoiding encountering a Covid-19 carrier. The overall density safety factor can also be coupled with other factors like reported infection rates to provide an enhanced user's situational representation.

The previously proposed architecture required the necessity of the health center experts in providing the positive test data to the database and registering the Covid-19 carrier for tracking. To test the prototype, the requirement of the health center experts was circumvented, and the database was loaded with fictitious Covid-19 carriers with actual locations against which the actual devices were tried to ensure they are proactively notified to avoid coming in contact. The current prototype implementation provides a more privacy enhanced solution by circumventing the necessity of the health center interference of user's medical status as it pertains to Covid-19 and their personal risks. Thus, it increases users' privacy by eliminating the direct linkage of users' medical records and their current Covid-19 test results and vaccine status as well as facilitate the usability of the application's rapid prototype. Figure 1 below shows an adapted architecture of the previously proposed centralized architecture that is used for the prototype.

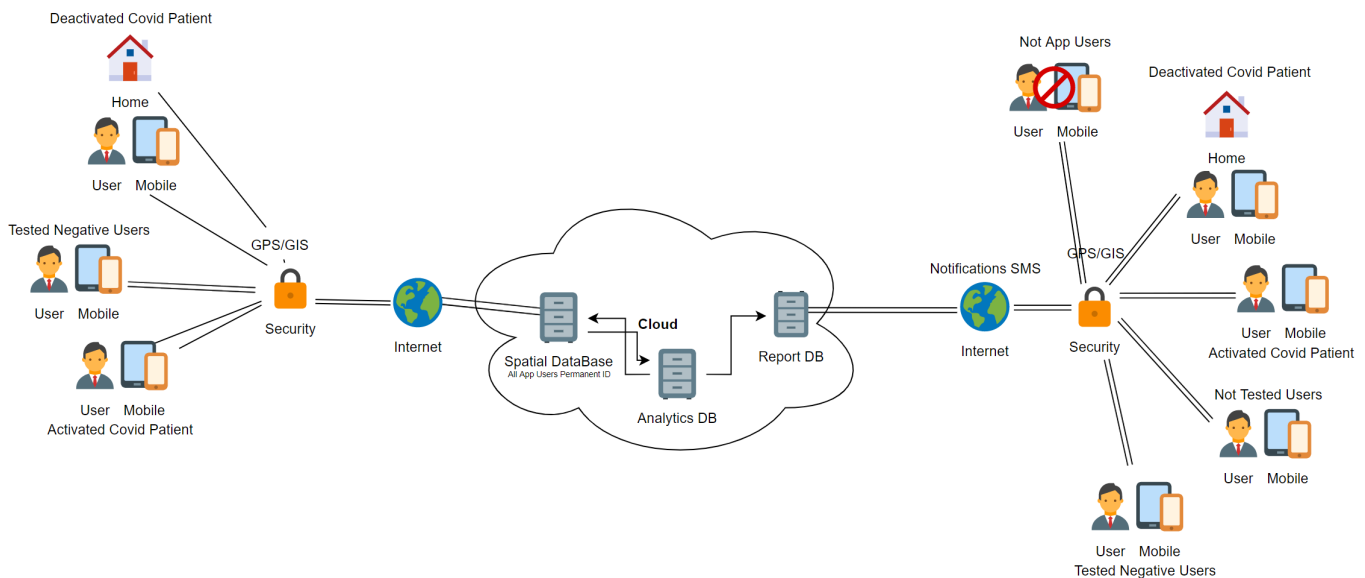


Figure 1: A diagram of the adapted architecture of the COVIDFREE centralized architecture prototype created using flat-color-icons.xml from the app.diagrams.net website.

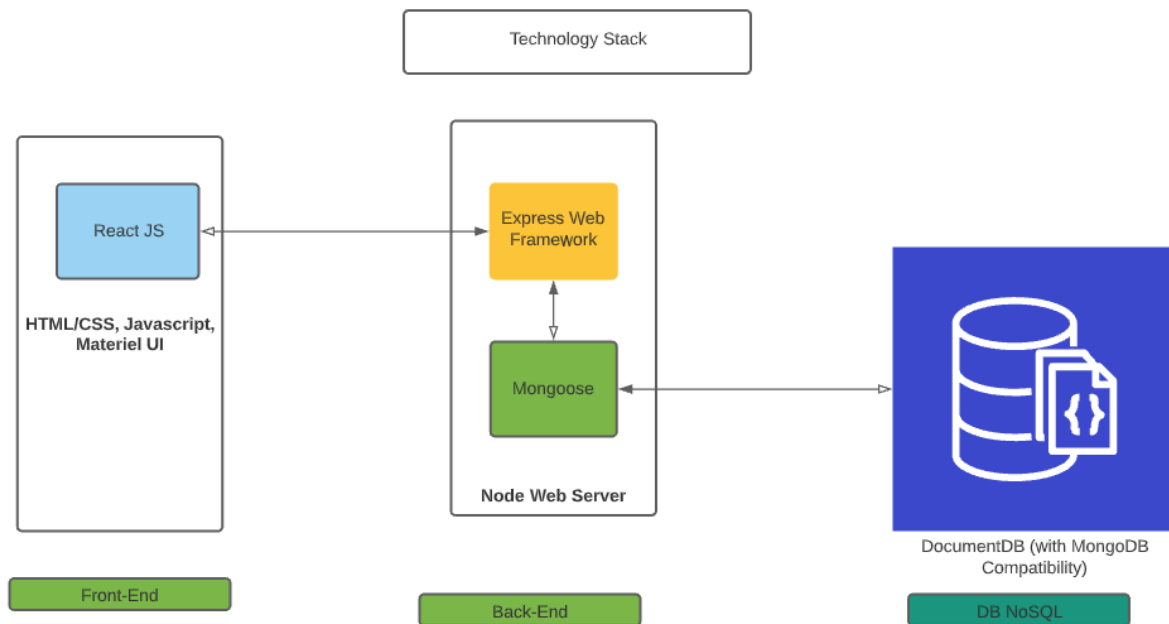


Figure 2: An illustration of the technology stack enabling the COVIDFREE application prototype created using flat-color-icons.xml from the app.diagrams.net website.

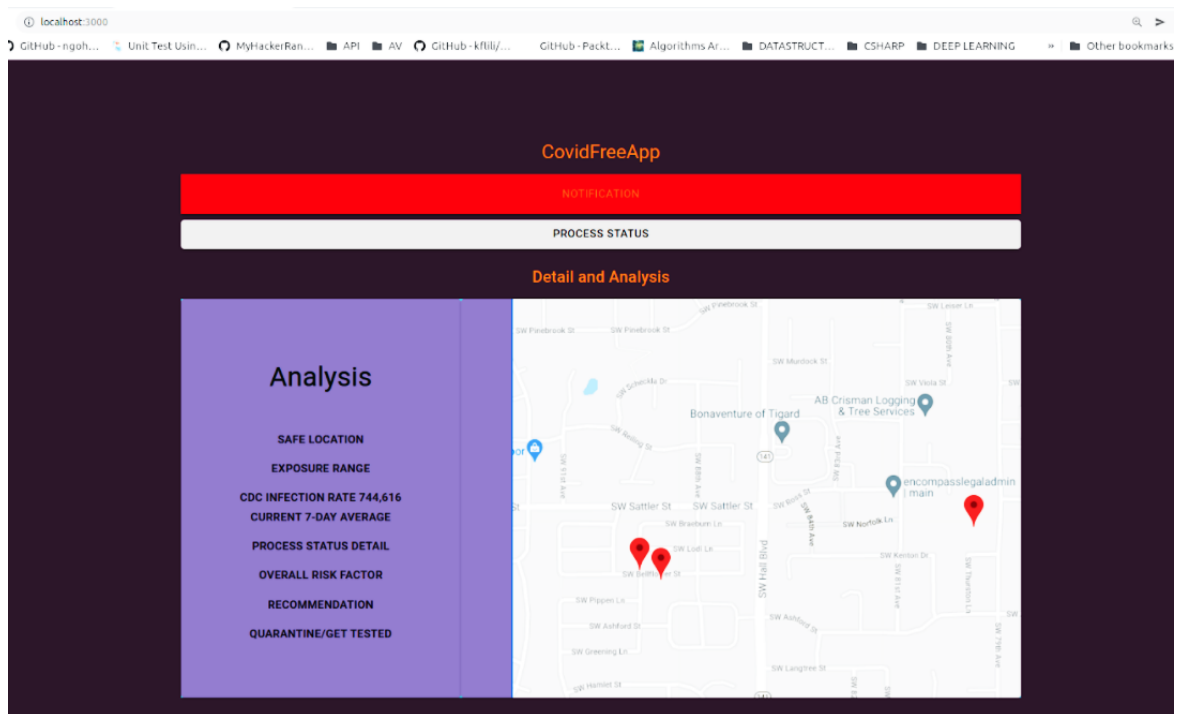


Figure 3: A screenshot showing the prototype of the COVIDFREE application's user interface

4. Used Technologies

To ensure the feasibility of the application in providing a prototype demo that other developers can use, many web application frameworks were considered for this project like Django, Ruby on Rails, MEAN and MERN frameworks. We decided to go with the MERN stack. MERN stands for MongoDB, Express, React, and Node, after the four key technologies that make up the stack. The technology stack enabling the prototype functionalities is shown in Figure 2 above.

The reason for choosing the MERN stack is that it provides an opportunity for rapid prototypes and proof of concepts. JavaScript is the primary programming language that is used in this project, and it offers the benefit of using one language for both the prototype's front-end and back-end. We used MongoDB to save the collected data. MongoDB is a document database model that maps to how developers think and code, and it provides a powerful, unified query API. MongoDB powers faster and more flexible applications. Node and Express are used for the back-end

implementation. Node is an asynchronous event-driven JavaScript runtime. It is designed to build scalable network applications.

5. App User Interface

Considering the front-end or the user interface UI, we used React. React is a free and open-source front-end JavaScript library for building user interfaces based on UI components.

We collect the data through the UI. Most data are collected dynamically without users' interaction. One of the most important data points that is collected is the user's location, which is associated with the user's registered phone number. We use the user's location to determine the safety of his position relative to nearby potential Covid-19 carriers while considering the infection rate in the user-specific location to determine the proper notification distance range.

The application's algorithm 1 shown below computes the best-case scenario for the end-user to avoid getting contaminated in the first place.

5.1. Algorithm

The application's algorithm as implemented in the code is designed with the virus prevention approach in mind. The source code is accessible on GitHub at link¹. The prototype UI is shown Figure 3.

Algorithm 1: User's Covid-19 Prevention Situational Response

Result: Action Recommendation with Warnings

```

Get User Safe Location;
Get User Current Position;
Get Users' Geolocation in Same Zip Code;
while User Not in Safe Location do
    Get Closest Covid-19 Carrier Position;
    Calculate Social Distance;
    if Social Distance greater than 100 Feet
    then
        | GREEN Notification;
    else if Social Distance greater than 50 Feet
    then
        | ORANGE Notification;
    else if Social Distance less than 50 Feet
    then
        | RED Notification;
    else
        | Break;
    end
end
end
    
```

6. App Demo and Use Cases

The prototype explores four different scenarios as use cases to ensure proactive notification to users to enable enough time for users to take appropriate actions based on their specific scenarios.

6.1. User in Safe Location

The first scenario involves the user being in their safe location or within 25 feet of the safe location. When this is the case, the user

is deactivated and is no longer tracked. The user is considered safe and providing no apparent threats to anyone regardless of their Covid-19 status since the safe location is also considered the quarantine location where the user is in isolation from the outside world. The user interface of this scenario shows the information and the notification that the users get in real time by having access to the app.

The notification button is triggered to a grey color notification, and only the user's home or safe location is shown on the map as the user is not in motion. The user can access all the analysis detail on what the process status means regarding all the notifications in detail by using the appropriate provided buttons on the application's analysis screen. The COVIDFREE application in this case recommends the user to avoid unnecessary trips to ensure they remain safe. The screenshot in Figure 4 below shows a representation of what the COVIDFREE application looks like with test demo data for this use case.

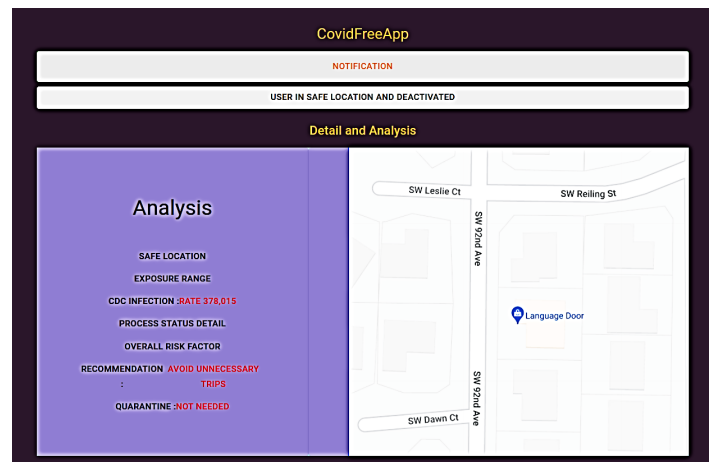


Figure 4: A screenshot showing the COVIDFREE application UI for a user in their safe location.

6.2. User out of Safe Location

The second scenario involves the user leaving their safe location and entering a state of motion moving from one location to another. When this is the case, the user is activated on the map and their position is tracked relative to his safe location and other potential Covid-19 threats agents or carriers. The user's test condition is considered when they are in motion relative to the potential Covid-19 carriers as the user runs a risk of interacting with a potential carrier. If the user is also a Covid-19 carrier then anonymously their location is also being flagged to be avoided by other users of the application providing a safe environment for all the application's users. The COVIDFREE application's user interface of this scenario shows the information and the notification that the users get in real time by having access to the app. The notification button is triggered to a green color notification if the user has not been in a vicinity of a Covid-19 carrier.

In this view, the user's current location and his safe location can be seen on the map interface of the application indicating the user is in motion and out of their safe location, and thus the user is

¹ <https://github.com/Unitercity2021/Covid-free-app>
www.astesj.com

active. The user can access all the analysis detail on what the process status means regarding all the notifications in detail by using the appropriate provided buttons on the application's analysis screen. The COVIDFREE application in this case recommends and reminds the user to social distance to ensure they remain safe throughout their trips. The screenshot in Figure 5 below shows a representation of what the COVIDFREE application looks like with test demo data for this use case.

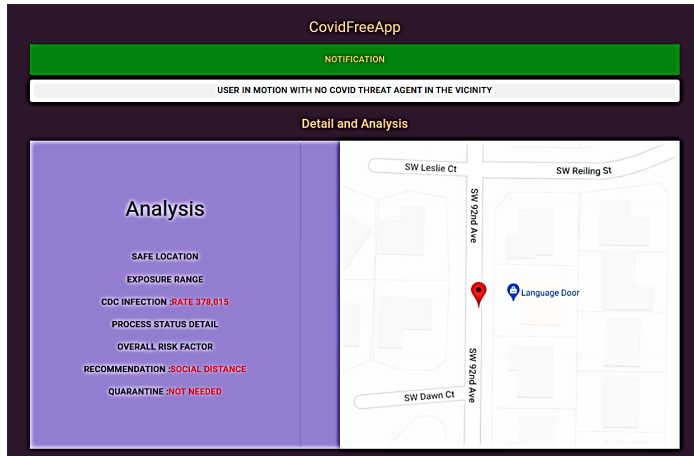


Figure 5: A screenshot showing the COVIDFREE application UI for a user outside their safe location and in motion, approaching no potential Covid-19 threat agents within their 100 ft radius.

6.3. User in motion with Covid-19 threat agents far away

The third scenario involves the user in a state of motion moving from one location to another and coming in the vicinity of an active Covid-19 carrier who is at a relatively far away distance of more than 50 feet from the user. When this is the case, the user remains activated on the map and their position is tracked relative to their safe location and the position of the potential Covid-19 threat agents or carriers. The user's Covid-19 test condition status is considered when they are in motion relative to the potential Covid-19 carriers as the user runs a risk of interacting with a potential Covid-19 carrier.

If the user is also a Covid-19 carrier then anonymously their location is also being flagged to be avoided by other users of the application providing a safe environment for all the application's users. The COVIDFREE application's user interface of this scenario shows the information and the notification that the users get in real time by having access to the app. The notification button is triggered to a yellow color notification since the user is in the vicinity of a Covid-19 carrier that is within their 100 feet radius but more than 50 feet away from them.

In this view, the user's current location and the Covid-19 carrier locations are shown on the map for the demonstration's sake and his safe location can also be seen on the map interface of the application when zoomed out indicating both that the user is in motion and out of their safe location as well as in the vicinity of Covid-19 threat carriers. The user can access all the analysis detail on what the process status means regarding all the notifications in detail by using the appropriate provided buttons on the application's analysis screen. The COVIDFREE application in this case recommends and reminds the user to consider quarantining when they return to their safe location. The screenshot in Figure 6

below shows a representation of what the COVIDFREE application looks like with test demo data for this use case.

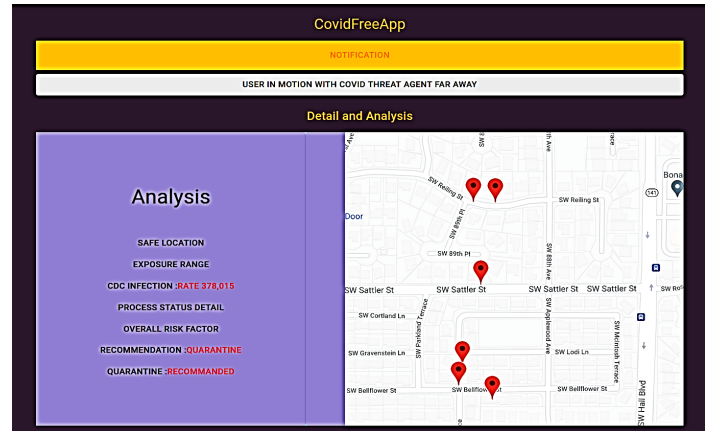


Figure 6: A screenshot showing the COVIDFREE application UI for a user outside their safe location and in motion, approaching potential Covid-19 threat agents that are relatively far away (more than 50 ft apart).

6.4. User in motion with Covid-19 threat agents nearby

The fourth scenario involves the user in a state of motion moving from one location to another coming in the vicinity of an active Covid-19 carrier who is relatively close by the user. When this is the case, the user remains activated on the map and their position is tracked relative to their safe location and potential Covid-19 threat agents or carriers. The user's test condition is considered when they are in motion relative to the potential Covid-19 carriers as the user runs a risk of interacting with a potential carrier. If the user is also a Covid-19 carrier then anonymously their location is also being flagged to be avoided by other users of the application providing a safe environment for all the application's users. The COVIDFREE application's user interface of this scenario shows the information and the notification that the users get in real time by having access to the app.

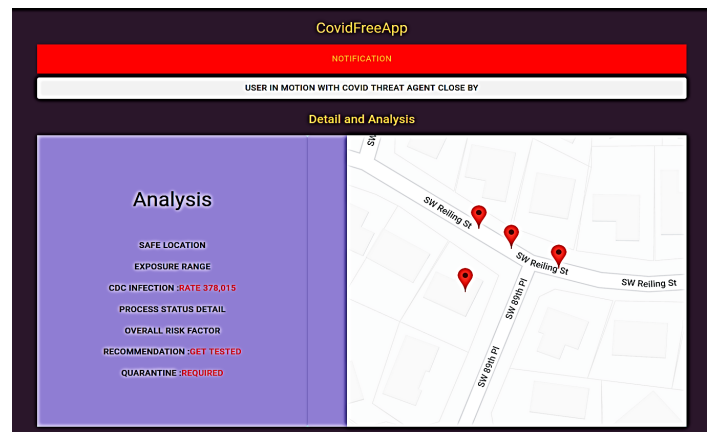


Figure 7: A screenshot showing the COVIDFREE application UI for a user outside their safe location and in motion, approaching potential Covid-19 threat agents that are relatively nearby (Less than 50 ft apart).

The notification button is triggered to a red color notification since the user is in the vicinity of a Covid-19 carrier that is relatively nearby. In this view, the user's current location and the Covid-19 carrier's location is shown on the map for demonstration's sake and their safe location can also be seen on

the map user interface of the application when zoomed out indicating both that the user is in motion and out of their safe location as well as in the vicinity of a Covid-19 threat carrier. The user can access all the analysis detail on what the process status means regarding all the notifications in detail by using the appropriate provided buttons on the application's analysis screen. The COVIDFREE application in this case recommends and reminds the user to consider getting tested and quarantining when they return to their safe location. The screenshot in Figure 7 below shows a representation of what the COVIDFREE application looks like with test demo data for this use case.

7. Conclusion

The current prototype presented in this paper demonstrates that the proof of concept for proactive notifications is feasible and can be further enhanced to provide efficient ways to process the users' data and provide swift notification to users while collecting the minimum amount of data from the users and providing multiple layers of abstraction to ensure security and privacy of the individuals and data quickly and optimally. To counteract the propagation of Covid-19 and increase citizens' safety and peace of mind, this article demonstrates the feasibility of COVIDFREE APP by creating a working prototype as a complementary and proactive technological solution that can minimize the likelihood of citizens contracting the Covid-19 virus.

The prototype application employs the previously proposed centralized architecture design to assist users in making educated decisions about how to comfortably and safely navigate from one location to another as well as when they can safely leave areas of isolation (such as their homes) and their immediate social groups. The prototype achieved the goal of improving users' situational alertness of high-risk sites around them. With better situational mindfulness, it is expected that users will likely feel more convinced and secure about their conduct, flexibility, and travel plans.

We hope this work stimulates parallel efforts to guarantee citizens leverage the available technologies to advance the citizen's safety and security, and eventually, to save citizens' lives. There are some opportunities for further developments. Our model gives a unique framework that is easy to use and configure for different machine learning models. For instance, one can implement a federated learning [7] model to optimize an independent user situational model and further enhance the user safety. As the application offers a real-time user notification to prevent contracting the virus and ensures safety, it also ensures the integrity of the information, and the trustworthiness of the data can be accessed using the zero-knowledge proof technique [8] even though the zero-knowledge proof technique is complex and in its early research stage. Nevertheless, it would provide a great way to gain public trust due to its security and privacy awareness implications.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors thank Arizona State University and the National Science Foundation for their funding support under Grant No. 1828010.

References

- [1] E.M. Mimo, T. McDaniel, "COVIDFREE App: The User-Enabling Contact Prevention Application," in 2020 IEEE International Symposium on Technology and Society (ISTAS), IEEE: 452–456, 2020, doi:10.1109/ISTAS50296.2020.9462186.
- [2] S. Shieh-zadegan, N. Alaghemand, M. Fox, V. Venketaraman, "Analysis of the Delta Variant B.1.617.2 COVID-19," *Clinics and Practice*, **11**(4), 778–784, 2021, doi:10.3390/clinpract11040093.
- [3] J.A. Plante, B.M. Mitchell, K.S. Plante, K. Debbink, S.C. Weaver, V.D. Menachery, "The variant gambit: COVID-19's next move," *Cell Host & Microbe*, **29**(4), 508–515, 2021, doi:10.1016/j.chom.2021.02.020.
- [4] N. Ahmed, R.A. Michelin, W. Xue, S. Ruj, R. Malaney, S.S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, S.K. Jha, "A Survey of COVID-19 Contact Tracing Apps," *IEEE Access*, **8**, 134577–134601, 2020, doi:10.1109/ACCESS.2020.3010226.
- [5] K. Michael, R. Abbas, R.A. Calvo, G. Roussos, E. Scornavacca, S.F. Wamba, "Manufacturing Consent: The Modern Pandemic of Technosolutionism," *IEEE Transactions on Technology and Society*, **1**(2), 68–72, 2020, doi:10.1109/TTS.2020.2994381.
- [6] Y.-C. Wu, C.-S. Chen, Y.-J. Chan, "The outbreak of COVID-19: An overview," *Journal of the Chinese Medical Association*, **83**(3), 217–220, 2020, doi:10.1097/JCMA.0000000000000270.
- [7] H.B. McMahan, E. Moore, D. Ramage, S. Hampson, B.A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," 2016.
- [8] S. Grzonkowski, W. Zaremba, M. Zaremba, B. McDaniel, "Extending web applications with a lightweight zero knowledge proof authentication," in Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology - CSTST '08, ACM Press, New York, New York, USA: 65, 2008, doi:10.1145/1456223.1456241.

Leakage-abuse Attacks Against Forward Private Searchable Symmetric Encryption

Khosro Salmani^{*1}, Ken Barker²

¹Assistant Professor, Department of Mathematics and Computing, Mount Royal University, Calgary, AB T3E 6K6, Canada

²Professor, Department of Computer Science, University of Calgary, Calgary, AB T2N 1N4, Canada

ARTICLE INFO

Article history:

Received: 12 January, 2022

Accepted: 07 April, 2022

Online: 19 April, 2022

Keywords:

Dynamic SSE

Cloud security

Access pattern leakage

Search pattern leakage

Data privacy

Leakage-abuse attacks

ABSTRACT

Dynamic Searchable Symmetric Encryption (DSSE) methods address the problem of securely outsourcing\updating private data into a semi-trusted cloud server. Furthermore, Forward Privacy (FP) notion was introduced to limit data leakage and thwart the related attacks on DSSE approaches. FP schemes ensure previous search queries cannot be linked to future updates and newly added files. Since FP schemes use ephemeral search tokens and one-time use index entries, many scholars conclude that privacy attacks on traditional SSE schemes do not apply to SSE approaches that support forward privacy. However, to obtain efficiency, all FP approaches accept a certain level of data leakage, including access pattern leakage. Here, we introduce two new attacks on forward-private schemes. We demonstrate that it is still plausible to accurately unveil the search pattern by reversing the access pattern. Afterward, the attackers can exploit this information to uncover the search queries and consequently the documents. We also show that the traditional privacy attacks on SSE schemes are still applicable to schemes that support forward privacy. We then construct a new DSSE approach that supports parallelism and obfuscates the search and access pattern to thwart the introduced attacks. Our scheme is cost-efficient and provides secure search and update. Our performance analysis and security proof demonstrate our approach's practicality, efficiency, and security.

1 Introduction

Cloud service providers offer various services that attract users and encourage them to outsource their personal data to reduce maintenance costs and increase user satisfaction, convenience, and flexibility. Nevertheless, these services come at the cost of losing complete control over the outsourced data, which raises security and privacy concerns. Although encrypting data before uploading it into the cloud addresses privacy concerns, it suffers from low efficiency. Keyword search is an essential requirement in these systems which is not supported by traditional encryption schemes. A naive solution is to download and decrypt all the encrypted documents to search for a keyword. Obviously, this solution suffers from excessive communication overhead and is inefficient.

Hence, Searchable Symmetric Encryption (SSE) schemes [1]–[4] were introduced to tackle this challenge. In SSE schemes, a cloud can perform search queries on user's outsourced data while the queries, results, and data are encrypted. In the other words, SSE schemes address both privacy challenges and the searchability requirement. However, the early approaches only work for static data

which means that no additions, updates, and deletions are feasible in a low-cost and efficient manner after the setup phase (securing and storing data into the cloud). Later, researchers proposed Dynamic SSE (DSSE) approaches [5]–[8]. These schemes enable users to update\modify the outsourced corpus arbitrarily in addition to performing search queries.

Moreover, SSE schemes support Multi-keyword [4, 9] search or Boolean [7, 10, 11]. Boolean schemes search for a single keyword and returns all of the documents that contain the queried keyword. Alternately, Multi-keyword search supports multiple keywords search which prevents unacceptably coarse results and improves result accuracy. Moreover, some of the SSE schemes support ranked search [4, 9] which means the cloud returns most relevant files by ranking them based on their relevance to the query.

However, DSSE approaches leak sensitive information such as search and access pattern. These methods employ deterministic queries \search tokens, which allow a server to determine if multiple queries consist of the same keyword (*search pattern*). furthermore, the matching document identifiers (*access pattern*) will be leaked af-

*Corresponding Author: Khosro Salmani, B113F- 4825 Mt Royal Gate SW, Calgary, AB T3E 6K6, (+1) 403-440-6492 & ksalmani@mtroyal.ca

ter each query. Many scholars have shown [12]–[14] that revealing such important meta-data can be used to obtain critical information and even to expose underlying plaintext data. Note that, in theory it is possible to design SSE scheme with no information leakage using several cryptographic primitives such as oblivious RAM, homomorphic encryption, and secure two-party computation [15]–[17]. However, these methodologies suffer from excessive computation power, low efficiency, and large storage costs.

Moreover, in Dynamic SSE approaches, it is still feasible to link previous search and update requests to a file that is recently added. For example, if we add a new document to the corpus, the server can determine whether the new document contains the keyword that we searched for in the past. Moreover, the cloud can execute the queries on deleted documents. To tackle these challenges, researchers introduced *Backward* and *Forward privacy* [5]. Backward privacy guarantees the privacy of deleted documents while forward privacy guarantees the privacy of newly added documents. No *efficient* approach currently provides full backward privacy.

Forward-private (FP) approaches employ one-time use search tokens to preserve the newly added documents' privacy [7, 6]. This means the search token for a keyword changes after being used in a query. Moreover, server index entries are ephemeral, which means the user generates new encrypted index entries after each time of access. Hence, the server cannot track the queries. As a result, Scholars believe that privacy attacks on traditional SSE approaches do not apply to forward-private schemes. In particular, in [14] the author believes these features “*highlights the importance of forward privacy*”, and in [18] the author believes with forward privacy these attacks can be thwarted and prevented. Furthermore, applying these attacks will become a cumbersome task, considering that forward-private schemes are primarily *dynamic* and provide *update* functionalities (including *add* and *delete*). Therefore, monitoring and linking the queries turns significantly harder, if multiple update requests occur between two search queries.

Nevertheless, this paper extends work initially presented in the Eleventh ACM Conference on Data and Application Security and Privacy [10] and shows that it is still possible to reveal the documents and queries accurately. All DSSE approaches accept a certain level of leakage to achieve an acceptable level of performance\efficiency [2, 7, 13, 19]. Hence, the primary objective is to increase the performance as high as possible while decreasing the leakage. In particular, *access pattern* leakage is among the acceptable leakages [5]–[7] and, thus, one of the open challenges that has not been addressed among the forward-private and traditional approaches.

In this paper with introducing two attacks we show that it is possible to retrieve the search pattern with high accuracy that can be exploited by previous attacks to unveil the search tokens and consequently the documents in FP approaches. Our introduced attacks **reverse-analyze** (see Section 4) the access pattern to recover the search pattern. The first attack is applicable on the forward-private DSSE approaches that only provide “add” functionality such as [7]. We modified the the first attacked (based attack) and introduced the advanced attack which can invade the forward-private DSSE approaches that provides both “add” and “del” functionalities such as [18].

In this paper, a new forward-private DSSE approach is intro-

duced to tackle this problem. In contrary with other scheme, our approach hides and obfuscates the access and search pattern and employs non-deterministic search tokens. All these features thwart and prevent the introduced attacks in this paper and also previous privacy and security attacks. Particularly, our contributions are:

1. Defining two concrete attacks and demonstrating its potential threats and privacy\ security risks.
2. Tackling the access pattern leakage challenge in DSSE approaches with forward privacy with a novel method.
3. Constructing an forward private DSSE scheme that support search and update (add and delete) operation. Furthermore, our *efficient* approach supports *parallelism* which is an important efficiency factor [7].
4. Providing a security proof against adaptive adversaries which verify the privacy and security of our method.
5. Demonstrating the efficiency of our approach in real-world by implementing it using real-world datasets.

The rest of this article is organized as follows. We present related work and the state-of-the-art work in Section 2. We then state the preliminaries in Section 3. In Section 4, we introduce two new attacks on current forward-private DSSE schemes, and in Section 5 we construct a new scheme that prevents the introduced attacks in Section 4. Experiments and evaluation are detailed in Section 6, and the security proof is provided in Section 7. Finally, we conclude our paper in Section 8.

2 Related Work

During the last decade, various privacy constructions and security definitions have been proposed for searchable encryption. Efficiency has always been a key requirement and a primary challenge in this research area. For example, Oblivious RAM [15] achieves full privacy and security without leaking any information to the server, but it is impractical for real-life applications because of its excessive computation costs. Hence, several approaches were designed [1, 2, 4, 9, 19] that selectively leak information (*e.g.*, search and access pattern). This means, these schemes accept a low level of information leakage to gain a higher level of efficiency.

Searchable Symmetric Encryption (SSE) and Public-key Encryption with Keyword Search (PEKS) are the two main divisions of searchable encryption schemes. In [20], the author introduced the notion of the public key encryption with keyword search, which followed by several methods [21]–[23] to improve the system cost and efficiency of PEKS approaches. In particular, these methods use one key for encryption and another key for decryption. Hence, only data users who possess the private-key can search the encrypted outsourced data. In this paper we focus on the SSE schemes and our introduced construction is build on symmetric security primitives.

The first SSE scheme was introduced by [1]. They employed a two-layered encryption to encrypt each keyword. However, they suggest a sequential search which impacts the search time (makes it linear to the document size). Later, in [2] the author used a secure

index structure called Bloom filters to address this issue. In [19], the author proposed a scheme which preserves the security of the outsourced data against an adaptive adversary. However, this comes at the cost of higher communication overhead and requires more memory space on the server side.

Nevertheless, traditional SSE approaches provide exact keyword search and cannot tolerate any imperfections or format inconsistency. In [24], the author addressed this issue and proposed a method in which resultant documents are selected based on the keyword similarity and closest possible matching documents. In [25], the author tackled the same challenge and proposes a scheme that decreases system cost and provides more efficiency.

Moreover, SSE schemes support Boolean [7, 10, 11] or Multi-keyword [4, 9] search. Boolean schemes search for a single keyword and returns all of the files that contain the respective keyword. On the other hand, multi-keyword ranked search solutions [4, 8, 26, 27] enhance the result accuracy by supporting multiple keywords search. In [4], the author introduced the notion of “coordinate matching” which is a similarity measure that matches as many keywords as possible. They also constructed a multi-keyword search approach using coordinate matching. However, previous methods only support single data owner. In [26], the author designed a new scheme with a trusted proxy that supports multiple data owners. In [8], [27] the author considered a system model with semi-honest cloud server and proposed verifiable SSE approaches that can detect a malicious server. Moreover, in [9] the author propose a multi-keyword ranked search scheme that solve the problem of search pattern, and co-occurrence information leakage. They introduce a novel chaining encryption notation which prevent the aforementioned information leakages.

Dynamic Searchable Symmetric Encryption (DSSE) methods were introduced to support *add*, *delete*, and *update* operations in an efficient manner. In particular, the author in [28] introduced a DSSE approach that preserve users’ privacy and security against adaptive chosen keyword attacks. In [29], the author proposed a new DSSE method called “Blind Storage” that hinders leaking sensitive information such as the size and number of stored documents. DSSE approaches employ interactive protocols which results in leaking more information about the outsourced data in compare with traditional SSE approaches. In [5], the author introduced the notion of forward-privacy and designed a forward private DSSE construction to address this issue. However, their proposed method suffers from low efficiency. In [6], the author improved the system efficiency by using trapdoor permutations and designed a more efficient forward private DSSE scheme. However, these approaches use sequential scan to execute a query which makes palatalization impossible. In [7], the author addressed this issue with designing a new forward private DSSE method that provides parallelism by design.

3 Problem Formulation

3.1 Preliminaries

For a finite set X , we employ $x \leftarrow X$ to represent that x is sampled uniformly from the set X . λ is the security parameter, and \parallel shows concatenation. Function $\text{neg}(k) : \mathbb{N} \rightarrow [0, 1]$ is negligible

if for all positive polynomial p , there exists a constant c such that: $\forall k > c, \text{neg}(k) < 1/p(k)$.

Definition 1 (Symmetric-key Encryption). A symmetric encryption scheme is a set of three probabilistic polynomial time (PPT) algorithms $\text{SE} = (\text{Gen}, \text{Enc}, \text{Dec})$ such that Gen takes an unary security parameter λ and generates a secret key k ; Enc takes a key k and n -bit message m and returns a ciphertext c ; Dec takes in a key k and a ciphertext c , and returns m if k was the key under which c was generated. The SE is required to be secure against chosen plaintext attack (CPA). We refer to [19] for formal definitions.

Definition 2 (Pseudorandom function). Let $F : \{0, 1\}^l \times \{0, 1\}^l \rightarrow \{0, 1\}^l$ be a deterministic function which maps l -bit strings to l' -bit strings. We define $F_s(x) = F(s, x)$ as a pseudorandom function (PRF) if: \forall PPT distinguishers $\mathcal{D} : |\Pr[\mathcal{D}^{F_s(\cdot)}(1^\lambda) = 1] - \Pr[\mathcal{D}^{f(\cdot)}(1^\lambda) = 1]| \leq \text{neg}(\lambda)$, where $f(\cdot)$ is a truly random function, and λ is the security parameter.

In Definition 3, the notation $(c_{out}, s_{out}) \leftarrow \text{protocol}(c_{in}, s_{in})$ denotes an interaction between client and server where c_{in} and s_{in} are the client and server input, and the c_{out} and s_{out} are the output of client and server after performing a protocol.

Definition 3 (DSSE Scheme). Let $D = \{D_1, \dots, D_n\}$ be a corpus of n documents, a Dynamic Searchable Symmetric Encryption consists of five PPT algorithms:

- $(sk, \perp) \leftarrow \text{GenKey}(1^\lambda, 1^\lambda)$: In this algorithm, the data owner (client) generates a secret key sk using the security parameter λ .
- $(I_c, I_s) \leftarrow \text{BuildIndex}((sk, D), \perp)$: In this algorithm the client’s secret key sk and document collection D are used to produce a client-index I_c , and server outputs index I_s .
- $(\perp, C) \leftarrow \text{Encryption}((sk, D), \perp)$: The client inputs secret key sk , and document collection D , and outputs the encrypted corpus $C = \{C_1, \dots, C_n\}$.
- $((I'_c, D_w), I'_s) \leftarrow \text{Search}((sk, I_c, w), (I_s, C))$: In this algorithm the client inputs the secret key sk , index I_c , and query w ; and it outputs the updated index I'_c , and resultant documents D_w . The server also, inputs the index I_s , and the encrypted document collection C and outputs the updated index I'_s .
- $(I'_c, (I'_s, C')) \leftarrow \text{Update}((sk, I_c, \text{op}, \cdot, \text{in}), (I_s, C))$: In this algorithm the client inputs the secret key sk , index I_c , and an operation $\text{op} = \text{add}$ or $\text{op} = \text{del}$, and an input “in”, which is parsed as a set of keywords w_{in} and a document identifier id_{in} . It outputs the updated index I'_c . The server inputs the index I_s , and the encrypted document collection C ; it outputs the updated index I'_s and updated encrypted document collection C' .

We call the first three protocols (GenKey , BuildIndex , Encryption) the **Setup phase**.

3.2 Our System Architecture

Our system architecture, as illustrated in Figure ??, consists of two parties: a cloud server and a client (data owner - user). The client is the actual owner of the data and intends to outsource its personal corpus into a cloud server for several reasons including maintenance costs. The client first creates an inverted index for each keyword.

Each entry in this index maps the respective keyword, w_i , to the documents IDs that contain w_i . The client then encrypts the documents and index entries and outsources them into the cloud server. Once the cloud receives a search request, it performs the query using the provided index and outputs the resultant files. Note that the documents and index entries are all encrypted, so the cloud server will not know the content of search tokens or the documents. Nevertheless, in see Section 3.5 we explain that like other related work [5]–[8], some meta-data may leak over time and after executing a number of queries.

3.3 Threat Model

In our approach, the server follows the prescribed protocol, however, it is keen to gather meta-data and information about the client. This type of cloud server is called *honest-but-curious* and is employed in many related work such as [5]–[7]. In addition, we suppose the server knows the encrypted index, documents, queries, and the employed encryption scheme, but it does not know the secret key.

3.4 A short overview of our approach

The client initiate the protocol by extracting keywords, $\Delta = \{w_1, w_2, \dots, w_m\}$, and creating the inverted plain-index. Each entry (id_i, L) in the index is a pair of an id_i and a list of L . Each keyword, w_i in the corpus corresponds to an id_i in the index, and L consists of all the files that contain w_i . To achieve our primary objectives which are hiding and obfuscating the access and search pattern, we inject random files IDs (noise) among the nodes in each list. The client is the only party who can distinguish the noise nodes. To monitor the lists, the user must keep a small index, \mathcal{I}_c (see Section 5.2) on her side. Then, the index entries and files will be encrypted and transferred to the cloud server. Upon receiving the data, the server stores the encrypted index entries, \mathcal{I}_s , and the encrypted corpus C and stands by for the first search or update request. Every query in our approach, \mathbf{q} , consists of a limited number of sub-queries $\mathbf{q} = \{q_1, \dots, q_k\}$. The fake/noise sub-queries are added to hide and obfuscate the search and access pattern. Once a query is received, $\mathbf{q} = \{q_1, \dots, q_k\}$, the server performs the sub-queries one-by-one, or parallelly if we employ the parallel algorithm, and returns the results. The user can retrieve the real results and discard the noise. Lastly, using new keys and IDs, new encrypted index entries will be created and sent to the cloud server. Note that the user (\mathcal{I}_c) and cloud (\mathcal{I}_s) indexes will be updated respectively.

3.5 Security Definitions

To gain efficiency, most of the SSE schemes leak some meta-data such as number of keywords, file size, and file IDs [4, 5, 19]. In addition, more meta-data may leak after performing each query. Thus, we start this sections by defining the leakage functions that show the leaked meta-data to the cloud server after executing each step of the protocol.

Definition 4 (Search pattern). Let $Q = (\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_t)$ be the query list over t queries. The search pattern over a query list Q is a tuple,

$SP = (\hat{w}_1, \hat{w}_2, \dots, \hat{w}_t)$, where \hat{w}_i , $1 < i < t$ is the encrypted keyword (or its hash) in the i -th query.

Definition 5 (Access pattern). The access pattern over a query list Q is a set, $AP = (R(\mathbf{q}_1), R(\mathbf{q}_2), \dots, R(\mathbf{q}_t))$ over t queries, where $R(\mathbf{q}_i)$, $1 < i < t$ is the i -th query's resultant document identifiers (result set).

To demonstrate the leakage to the server, we employ leakage function \mathcal{L}^{op} which indicates the information revealed to the adversary after executing operation op . We first define the $\mathcal{L}^{\text{Setup}}$ and $\mathcal{L}^{\text{Search}}$, and then demonstrate the information leakage of Update through Definition 6.

- $\mathcal{L}^{\text{Setup}}(D) = \{N, n, (id(D_i), |D_i|, C_i)_{1 \leq i \leq n}\}$, where N is the number of server index entries, \mathcal{I}_s ; n is the number of documents, $id(D_i)$ is the document D_i 's identifier, $|D_i|$ is the size of document D_i , and C_i is the encrypted corpus.
- $\mathcal{L}^{\text{Search}}(\mathbf{q}_i) = \{R(\mathbf{q}_i), |R(\mathbf{q}_i)|, C_{\mathbf{q}_i}\}$ where \mathbf{q}_i is a client's query, and $R(\mathbf{q}_i)$ is the resultant document identifiers, and $C_{\mathbf{q}_i}$ is the encrypted resultant documents.

Definition 6 (Forward privacy). A SSE scheme is forward private if the update leakage function $\mathcal{L}^{\text{Update}}$ is limited to: $\mathcal{L}^{\text{Update}}(in, D_i, \text{op}) = \{id_m(D_i), |w_m|, |D_i|, \text{op}\}$.

To define the security of the DSSE scheme we employ the standard simulation model which requires a real-ideal simulation [5, 7]:

Definition 7 (DSSE Security). Let $\text{DSSE} = (\text{GenKey}, \text{BuildIndex}, \text{Encryption}, \text{Search}, \text{Update})$ be a DSSE scheme. Let \mathcal{A} be an adversary (server), and the $\mathcal{L}^{\text{Setup}}$, $\mathcal{L}^{\text{Search}}$, and $\mathcal{L}^{\text{Update}}$ be the leakage functions. The following describes the real and ideal world:

- **Ideal** $_{\mathcal{F}, \mathcal{S}, \mathcal{Z}}^{\text{DSSE}}(\lambda)$: An environment \mathcal{Z} sends the client a set of documents D to be outsourced. The client forwards them to the ideal functionality \mathcal{F} . A simulator \mathcal{S} is given $\mathcal{L}^{\text{Setup}}$. Later, the environment \mathcal{Z} asks the client to run an Update or a Search protocol by providing the required information. The search request is accompanied with a keyword w . For an update request, \mathcal{Z} picks an operation from $\{\text{add}, \text{del}\}$. Add requests are accompanied with a new document and del requests contain a document identifier. The client prepares and sends the respective request to the ideal functionality \mathcal{F} . Using $\mathcal{L}^{\text{Update}}$ and $\mathcal{L}^{\text{Search}}$, \mathcal{F} notifies \mathcal{S} of leakages. \mathcal{S} sends \mathcal{F} either abort or continue. The ideal functionality \mathcal{F} sends the client either abort or "success" for Update, or set of matching document identifiers for Search. Finally, the environment \mathcal{Z} outputs a bit as the output of the experiment.

- **Real** $_{\Pi_{\mathcal{F}, \mathcal{A}, \mathcal{Z}(\lambda)}}^{\text{DSSE}}$: An environment \mathcal{Z} sends the client a set of documents D to be outsourced. Then, the client executes the $\text{GenKey}(1^\lambda)$ to generate the key sk and starts the BuildIndex and Encryption protocols with the real world adversary \mathcal{A} . Later, the environment \mathcal{Z} provides the required information and asks the client to run a Search or an Update request. The search request contains a keyword w to search for. \mathcal{Z} picks an operation from $\{\text{add}, \text{del}\}$ for an update request. Add requests are accompanied with a new document and del requests contain a document identifier. The client then executes

the real-world protocols with the server on the inputs that are selected by \mathcal{Z} . The client outputs either **abort** or “**success**” for **Update**, or a set of matching document ids for **Search**. \mathcal{Z} observes the output. Finally, outputs a bit b as the output of the experiment.

We say that a DSSE scheme (Π_F) emulates the ideal functionality \mathcal{F} in a semi-honest model, if for all PPT real world adversaries \mathcal{A} , there exists a PPT simulator \mathcal{S} such that for all polynomial-time environments \mathcal{Z} , there exists a negligible function $\text{negl}(\lambda)$ on the security parameter λ such that [5]:

$$|Pr[\mathbf{Real}_{\Pi_F, \mathcal{A}, \mathcal{Z}}^{\text{DSSE}} = 1] - Pr[\mathbf{Ideal}_{\mathcal{F}, \mathcal{A}, \mathcal{Z}}^{\text{DSSE}}(\lambda) = 1]| \leq \text{negl}(\lambda).$$

4 Attack methods

The first step of the attacker to launch an attack is to put the file IDs in a random order. For instance, $(id(D_5), id(D_7), id(D_4))$ is a valid order for a corpus with three documents $\{D_4, D_5, D_7\}$. Based on the chosen arbitrary order, the server creates a bit-string after executing each query. Each bit will be set to one if the corresponding file exists in the result set and to zero otherwise. For instance, suppose after executing a query, \mathbf{q}_i , the results set is the $R(\mathbf{q}_i) = \{D_4\}$. Thus, 001 is the corresponding bit-string that is generated based on result set of the current query. Moreover, to keep track of the frequency of each bit-string, the cloud server creates a Search Pattern Map (SPM) which is a hash map data structure. The attacker’s main challenge is to track the queries and since the search tokens are one-time use, storing them is pointless. However, the bit-strings that are created in our attacks can be employed as search token identifiers. Hence, the attacker stores them in the SPM along with the number of times that each token is searched.

In other words, the search tokens are ephemeral and change after each use, but the result set for each keyword remain the same and it becomes a major vulnerability for forward private schemes because of the access pattern leakage. For instance, $(11, \{001, 7\})$ can be possible element in SPM that demonstrates the keyword with 001 bit-string has queried seven times. The “11” number is the hash map key that starts from zero and increments by one after adding a new element. The complexity (number of elements) of the SPM is $O(m)$ where m is the number of keywords.

Like other related work [30], in our attacks it is assumed that the bit-strings are unique. To challenge this assumption, we extracted and studies 1927 keywords from the 50,000 files in Enron email dataset [31]. The results shows a scarce 0.2% conflict probability. In other words, there were only 2 conflicts among the investigated files. As a result, the search pattern can be recovered with 99.8% accuracy using our attack. In addition, remark that the keywords that had conflicts were among the very low frequency keywords, thereby, perhaps the cloud server is not interested in. Furthermore, the conflict probability significantly decreases as the number of files in the corpus increase. This is because the state space of bit-string’s, all possible bit strings set, expands and becomes larger. Nevertheless, we later address this attacker’s challenge and describe how keywords with unique bit-string can be distinguished. We emphasize that all assumption in related work [13, 14, 30, 32] and our work are consistent and we add no new assumption in this attack.

The basic attack is explained in Algorithm 1. Briefly, once each query is executed, the cloud server looks in the SPM to find a match for the resultant bit-string ($r_bitString$). If there is a match, the server increments the respective frequency by one, otherwise, the new bit-string will be added to SPM with frequency of one (line 24).

Algorithm 1 Basic Attack

input: SPM, $r_bitString$

output: updated SPM

```

1: found = false
2: for each  $e \in$  SPM & until !found do
3:    $e_{tmp} = e.bitString$ 
4:    $r_{tmp} = r\_bitString$ 
5:   flag = true
6:   while flag &  $e_{tmp} > 0$  do
7:      $e_{rem} = e_{tmp} \bmod 2$ 
8:      $r_{rem} = r_{tmp} \bmod 2$ 
9:     if  $e_{rem} \neq r_{rem}$  then
10:      flag = false
11:     end if
12:      $e_{tmp} / = 2$ 
13:      $r_{tmp} / = 2$ 
14:   end while
15:   if flag then
16:     found = true
17:     match =  $e$ 
18:   end if
19: end for
20: if found then
21:    $match.bitString = r\_bitString$ 
22:    $match.frequency++$ 
23: else ▷ new keyword found
24:   Add ( $r\_bitString$ , 1) to SPM
25: end if

```

Nevertheless, the length of the bit-string can be affected by “add” operation. In other words, adding a new file increases the length of the bit-string. To tackle this issue, the new file ID will be added to the left of the arbitrary order by the cloud server. Recall the previous example and suppose the server has received a new request to add D_6 to the dataset. The updated arbitrary order will be $(id(D_6), id(D_5), id(D_7), id(D_4))$. Hereafter, SPM bit-strings are a bit shorter than queries’ bit-strings. However, this does not stop the server\attacker from recovering the search pattern, because, the attack algorithm compare the SPM and query bit-strings bit-wisely starting from left bit to the right. The algorithm halts (line 6) when it achieves the last bit of the respective SPM bit-string. For instance, suppose $\{D_4\}$ and 001 are the result set and respective bit-string of query, \mathbf{q}_i . If the user issues the same query \mathbf{q}_i again, after adding D_6 and of course with a new search token, the resultant bit-string would be either 0001 or 1001. Remark that only one can happen at a time, because either the new file, in this case D_6 , contains respective keyword in \mathbf{q}_i or not. Hence, if the server detects a bit-string in SPM that matches the first three bits (from right-side) of the resultant

bit-string, it can be confident that these two token IDs refer to same keyword. Remark that, the respective bit-string will be updated to $n + 1$ from n -bit string in line 21 of the algorithm where n is the number of files. This means, in our example the bit-string will be updated to a four-bit from a 3-bit string.

The traditional attacks are not effective on schemes that support forward privacy. The main reason is that forward-private approaches hide and obfuscate the search pattern to a certain extent. However, by applying our attack on schemes with forward privacy, we reveal the search pattern. Once the attacker possess the search pattern, forward-private approaches will become susceptible against previous attacks. The output of our attack can be exploited by frequency-based attacks such as [12]–[14], [32]. Moreover, after applying a small modification, our attack can be used by occurrence-based attacks such as [30]. To support the occurrence-based attacks the attacker creates a $n \times m$ matrix \mathcal{M} instead of SPM. In this matrix each column represents a bit-string\keyword, and each row corresponds to a document. We set the value of an entry to zero if the respective keyword does not exist in the corresponding document, and to one otherwise. Once a query is executed, the cloud server updates the value of the respective entry, If it finds the same bit string, or it adds the bit-string as a new keyword otherwise. If a new file is added, the cloud server append a new row to the matrix \mathcal{M} .

To address the problem of distinguishing the keywords with the same result set, the cloud server can inject a limited numbers of documents into the corpus (keywords with the same result set). For instance, suppose $\{k_1, k_2\}$ and $\{k_3, k_4, k_5\}$ are two groups of keywords that have the same result set. The attacker can distinguish k_1 from k_2 by injecting a file that contains either of the keywords. The same method can be used to make the other group keywords distinguishable. To maximize the efficiency, we should minimize the number of injected files. Hence, the attacker creates new files that contains only one keyword from each group. For instance, the attacker creates a file that contains k_1 and k_3 from the first and second group. It also generates a another file which only contains k_4 . With injecting only two files these keywords will become distinguishable. Generally, suppose we have l groups of keywords that possess the same result set, $\{\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_l\}$, the minimum number of required injected documents is $\max_{j=1}^l \{|\mathcal{P}_j|\} - 1$, in which $|\mathcal{P}_j|$ shows the cardinality of j -th group.

This techniques is also employed in many related work such as [13, 32, 14] in which the cloud server sends the documents of its choice to the user. The client then encrypts and transfers them back to the cloud [14]. For example, consider a company that uses an automatic email process. Remark that in both occurrence-based and occurrence-based attacks the attacker commonly benefits from public information and auxiliary knowledge that rectifies the indistinguishable keywords challenge in advance. For instance, the attack in [12] benefits from public web facility services such as Google Trends[®].

Nevertheless, the basic attack is not applicable on DSSE approaches that provide “del” functionality. We modified the basic attack, see Algorithm 2, that can successfully attack DSSE schemes that provide both *add* and *del* functionality. In our *advanced attack*, a new bit-string, $d_bitString$, will be generated by the cloud server to monitor the deleted documents. Each bit in the $d_bitString$ represents a document in the corpus (considering the same arbitrary

order). Each bit will be set to zero if the respective file still exists in the corpus, and to one if it is deleted. After executing a delete request, the cloud updates this bit-string accordingly. Once a query is executed, the resultant bit-string will be bit-wisely compared to the data in SPM, but this time, we ignore the bits that their corresponding files are deleted. We demonstrate the changes with red lines in Algorithm 2.

Algorithm 2 Advanced Attack

input: SPM, $r_bitString$, $d_bitString$

output: updated SPM

```

1: found = false
2: for each  $e \in$  SPM & until !found do
3:    $e_{tmp} = e.bitString$ 
4:    $r_{tmp} = r\_bitString$ 
5:    $d_{tmp} = d\_bitString$ 
6:   flag = true
7:   while flag &  $e_{tmp} > 0$  do
8:      $d_{rem} = d_{tmp} \bmod 2$ 
9:     if ! $d_{rem}$  then
10:        $e_{rem} = e_{tmp} \bmod 2$ 
11:        $r_{rem} = r_{tmp} \bmod 2$ 
12:       if  $e_{rem} \neq r_{rem}$  then
13:         flag = false
14:       end if
15:        $e_{tmp} / = 2$ 
16:        $r_{tmp} / = 2$ 
17:     end if
18:      $d_{tmp} / = 2$ 
19:   end while
20:   if flag then
21:     found = true
22:     match =  $e$ 
23:   end if
24: end for
25: if found then
26:   match.bitString =  $r\_bitString$ 
27:   match.frequency++
28: else ▷ new keyword found
29:   Add ( $r\_bitString$ , 1) to SPM
30: end if

```

4.1 An example of our attack with “del” operation

Assume there are four documents, $\{D_1, D_2, D_3, D_4\}$, and four keywords, $\Delta = \{w_1, w_2, w_3, w_4\}$, in the user’s corpus. Moreover, $(id(D_4), id(D_3), id(D_2), id(D_1))$ is the arbitrary order that the attacker\cloud uses to create the bit-strings. Suppose D_1 contains $\{w_1, w_2, w_3\}$, D_2 includes $\{w_2, w_3, w_4\}$, D_3 has $\{w_1, w_3\}$, and D_4 contains $\{w_2, w_4\}$. Furthermore, we assume the setup phase is successfully executed, and the encrypted index and documents are outsourced.

Case 1: Searching for a keyword for the first time. The user searches for all documents that contain w_1 , so he generates an ephemeral search token (q_1) and send it to the server. Upon

receiving the search token, the server finds the related index entries and the respective encrypted documents ($R(\mathbf{q}_1) = \{D_1, D_3\}$). Since there is no bit-string in the SPM that matches the current bit-string, the server adds the respective bit-string (0101, 1) to the SPM (1 is the frequency). The user then generates and sends an encrypted query (\mathbf{q}_2) to the server to search for w_3 . After returning the results ($R(\mathbf{q}_2) = \{D_1, D_2, D_3\}$), the server adds (0111, 1) to the SPM.

Case 2: Searching again for a keyword that exists in the SPM. Now imagine the user searches again for w_1 with a new ephemeral search token (\mathbf{q}_3). Once the query executed, the server searches for the resultant bit-string (0101) in the SPM. Since the bit-string already exists in the SPM, the server only updates its frequency (0101, 2).

Case 3: Adding a new document. The user then adds a new document (D_5) that contains (w_1, w_2, w_4). The server updates the arbitrary order to $\{id(D_5), id(D_4), id(D_3), id(D_2), id(D_1)\}$ respectively. Now the user issues a new query (\mathbf{q}_4) to search for w_4 for the first time. The resultant bit-string will be 11010. Hence, (11010, 1) will be added to the SPM. Note that at this point the bit-strings in the SPM may have different length (4 and 5). The user may also search again for a keyword after adding a new document. For example, suppose the user searches again for w_3 . The new resultant bit-string is 00111. The server looks for a bit-string in the SPM that is either equal to our current bit-string or is equal to the first four bits (from right-side) of the resultant bit-string. In this case the server will find the (0111, 1) entry and will update it to (00111, 2) respectively.

Case 4: Deleting a document. To monitor the deleted documents, the attacker creates a bit-string, $d_bitString$, in which each bit represent a document and its value demonstrates whether the respective file is deleted (=1) or not. This vector will be updated after each delete request. Once a query is executed, the resultant bit-string will be bit-wisely compared to the data in SPM, but this time, we ignore the bits that their corresponding files are deleted.

Assume the user asks server to delete D_3 . The server deletes the document and updates the $d_bitString$ to 00100. If the user searches for w_1 again, the resultant bit-string will be 10*01 (“*” means its value is not important and can be 0 or 1). Since D_3 is deleted (based on the $d_bitString$), the server ignores the value of that position when it is searching for the current bit-string in the SPM.

5 Construction

To prevent the attacks that we introduced in the previous section, we build and construct a new dynamic SSE scheme that supports forward privacy. Our construction hides and obfuscates the access pattern to thwart the above privacy attacks. In our approach, the client first creates an inverted index for each keyword in the corpus. In an inverted index, every entry maps a keyword to the corresponding document IDs that contains the respective keyword. We add fake/noise document IDs among each keyword result-set to hide and obfuscate the access pattern. The fake IDs can only be distinguished by the client. In addition, in our approach each query \mathbf{q}_i consists of a limited number of sub-queries $\mathbf{q}_i = \{q_{i1}, q_{i2}, \dots, q_{ik}\}$. All sub-queries except one are noise which can only be recognized by the user and sub-query searches for a keyword. We propose two methodologies to inject the noise file IDs:

1. **Random injection.** We first determine a threshold, τ_d , that represents the lower and upper bound of noise injections in a specific result-set. Once τ_d is set, we arbitrarily inject file IDs into the result set of the every keyword.
2. **Aforethought injection.** Our main objective is to flatten the the number of times that each document accessed. With this strategy the number of times that each document is accessed will be in the same range as others. This makes it much harder for the attacker to gain information about the data, search tokens, and the files by using the access pattern meta-data. To achieve this objective, the client generates an Access Pattern Vector (APV) that stores the number of times that each file is accessed. For instance, $\langle 1, 0, 4, 2 \rangle$ demonstrates D_1, D_3 , and D_4 are accessed one, four, and two times since launching the system. This information is valuable in our approach and will help to choose the fake sub-queries wisely. For each query, the user monitor and analyze the APV vector and selects the keywords that are accessed less in compare with others. Considering our earlier example ($\langle 1, 0, 4, 2 \rangle$), the client can inject keywords that return D_1 and D_2 to straighten the access pattern vector.

Creating the noise sub-queries are a significant challenge in the aforethought injection. To query the less requested files in the corpus, the user must be able to identify the keywords that return a specific file ID in their resultant-set. The number of files in the corpus is myriad and increasing over time, thereby keeping this information on the users' devices is not a realistic approach. In addition, to hide the clients' foot tracks (activities), we intend to put each file ID into more than one keyword's result-set. This makes the above solution more infeasible.

To tackle this problem, we first label all of the files with a number in a random order. Note that even storing these labels on the users' machines are not practical. Hence, we then generate an arbitrary number called δ . Afterward, we begin with δ and label each document successively. remark the files are shuffled before being labeled with successive numbers, so, the labels will not leak additional meta-data about the files. To identify the keyword w that a specific file ID, D_i , is injected in, we employ the $G_w(\cdot)$ PRF. $G_w(\cdot)$ accepts a key k and a document label and returns a keyword number, m , where $m \in [1..m]$. For instance, in $G_n(k, 7654321) = 123$ the user is searching for a keyword number (123) that holds a document that is labeled as 7654321. In other words, after calculating $G_n(k, 7654321) = 123$, the user learns that the file with 7654321 label is injected in the w_{123} ' result set.

To inject each file ID a number of times, we define a new parameter called step, s_δ . Step (s_δ) shows the number of result-sets that each file ID is injected into. As a result, we increase the label number by s_δ instead of labeling them consecutively. By exploiting this technique we expedite the flattening process of the APV vector. Therefore, instead of possessing one label, each file will have s_δ labels that is reserved for the respective file. For instance, suppose $s_\delta = 3$ and we want to search for keyword w_i that returns the document that is labeled as 7654321. Hence, the client executes $G_n(k, 7654321) = 123$, $G_n(k, 7654322) = 77$, and $G_n(k, 7654323) = 2105$. This means, keywords w_{123} , w_{77} , and w_{2105} are having file 7654321 in their result-sets. Consider that, due to the

definition of the pseudo random function the keyword numbers are not necessarily successive even if the the input labels are.

5.1 Our linked-list structure

In this section we explain a customized linked-list data structure that we employ in our scheme. Our linked-list consists of 4 tuples: (*id*, *type*, *data*, *next*). In particular, *id* refers to the ID of a node, *type* shows whether a node is real *R* fake\noise (F), *data* is a file identifier, and *next* refers to the ID of the next node in the respective linked-list. The red elements will be secured and encrypted using the user's key, while an ephemeral key will be used to encrypt the orange elements. Hence, *type* will be encrypted using the user's key and, we will be using an ephemeral key to encrypt *data* and *next*. Note that elements in black (*i.e.*, *id*) is plaintext data. To prevent leaking any additional meta-data, we use a random even number for real and a random odd number for noise nodes. In addition we set the *next* to null if a node does not have a successive node. Consider that, the ephemeral key will be provided for the server if the respective keyword is being queried, but we never share the user's secret key with server. Hence the server will never know which nodes are fake and which ones are real.

- **AddNode**($\mathbb{L}, k_1, k_2, id, type, data$): This function employs the input data to append a node to the beginning of the current linked-list, \mathbb{L} .
- **RestoreList**(k, id_h): This function looks for the node with the provided ID; it then retrieves *type* and decrypts *next* and *data* and looks for the next node in the linked-list. The process stops when the algorithm reaches the last node in the linked-list (*i.e.*, *next* = null).

In our approach, all of the secure inverted index is constructed using the aforementioned linked-list data structure. To add more security, the client injects fake\noise nodes in each and every linked-list to hide and obfuscate the relevance between a keyword, w_i , and its corresponding linked-list, L . Note that the noise will not be injected if it already exists in the linked-list. for instance, assume that one of the inverted index entries is ($w_7, \{D_3, D_6, D_5\}$). The objective is to inject D_6 and D_9 in random positions in the respective linked-list. however, we only inject D_9 because D_6 is already in the linked-list. Now suppose after injecting the noise node(s) the index entries will be ($w_7, \{D_3, D_6, D_9, D_5\}$). Hence, the user generates four nodes ($id_1, R, id(D_3), id_2$), ($id_2, R, id(D_6), id_3$), ($id_3, F, id(D_9), id_4$) ($id_4, R, id(D_5), null$). Consider that, since the nodes are encrypted and will be sent in a random order to the server, the attacker is not able to link them.

5.2 Our scheme

In this section we demonstrate and describe how each algorithm in Definition 3 operates:

- **GenKey**: let $SE = (Gen, Enc, Dec)$ be a CPA-secure symmetric encryption scheme. Let $G_n(\cdot)$, $G_{id}(\cdot)$ and $G_w(\cdot)$ be three PRFs and $GenPK(1^\lambda)$ be a key generator function. The following describes (sk, \perp) $\leftarrow GenKey(1^\lambda, 1^\lambda)$:
 - 1: $k_{SE} = SE.Gen(1^\lambda)$

- 2: $k_G \leftarrow GenPK(1^\lambda)$
- 3: *return* $sk = (k_{SE}, k_G)$

We employ a secret key, sk , to fulfill the encryption objectives. sk is a tuple of two, k_{SE} and k_G . The former will be used to encrypt the documents and latter for $G_w(\cdot)$, $G_n(\cdot)$ and $G_{id}(\cdot)$ functions.

Algorithm 3 ($\mathcal{I}_c, \mathcal{I}_s \leftarrow BuildIndex((sk, D, s_\delta), \perp)$)

Client

- 1: $\Delta = ExtractKeywords(D)$
 - 2: $\delta = Rand()$
 - 3: $lbl_{next} = \delta$
 - 4: $D' = \{\}$
 - 5: **while** $D \neq empty$ **do**
 - 6: $D_{cur} =$ randomly choose one doc and assign lbl_{next} to it
 - 7: $D' \cup D_{cur}$
 - 8: $id_{next} + = s_\delta$
 - 9: **end while**
 - 10: $PI = BuildPlainIndex(\Delta, D', s_\delta)$
 - 11: Create \mathcal{I}_c and APV and initialize all elements to zero
 - 12: $\mathbb{L} = GenLinkedList(sk, PI)$
 - 13: Send \mathbb{L} to server
- Server**
- 14: Generate \mathcal{I}_s using \mathbb{L}
-

- **BuildIndex**. In this algorithm (see Algorithm 3), after extracting the keywords, $\Delta = \{w_1, \dots, w_m\}$, we assign an label\id to each file according to the value of the δ (the starting point), and s_δ (step). Afterward, the client index \mathcal{I}_c , and its corresponding linked-lists will be created. To enable the client to generate the search queries, the client must store a $m \times 2$ look-up table (index), \mathcal{I}_c . In particular, this table stores length of the list, len_i and the number of nodes, cnt_i , that is generated for each keyword. Moreover, the APV (access pattern vector) will be created and initialized to zero. Once the index entries are generated, they will be sent to the cloud server. Note that a random number will be assigned to δ which is generated by $Rand()$.

We explain generating the plain-text inverted index PI in Algorithm 4. Using the aforethought injection method, we first generates the noise nodes (line 5-9), and then we append the real nodes to their corresponding list (line 11-15). Note that as we mentioned in Section 5.1, every entry in PI consists of two tuples which are a keyword and the receptive list (w_i, L). recall that beside the file label, each node in the list keeps a type (F\R). For instance, D_2 is fake, and D_4 and D_8 are real nodes in ($w_{23}, \{\{D_8, R\}, \{D_2, F\}, \{D_4, R\}\}$). In addition, to decide the lists that each file ID must be injected in, we employ $G_w(k_G, doc_{id})$ function. Consider that this is process happen in the setup phase and only for once. Next, a linked-list will be generated for each generated list in the previous step. The node IDs are one-time use and will be generated by the G_{id} function. In particular, the G_{id} function uses a counter, ctn , which shows the number of nodes that are created for the respective keyword, w_i . The client store this number in the client index ($\mathcal{I}_c[i][0]$).

Algorithm 4 BuildPlainIndex

```

1: procedure BuildPlainIndex( $\Delta, D', s_\delta$ )
2:    $PI = \{\}$ 
3:   for all  $w_i$  in  $\Delta$  creates an empty  $L_i$ 
4:   for all  $D_j$  in  $D'$  do
5:     for  $k = 0$  to  $s_\delta - 1$  do
6:        $i = G_w(k_G, (lbl(D_j) + k))$ 
7:       if  $D_j \notin L_i$  then
8:          $L_i \cup \{D_j, F\}$ 
9:       end if
10:    end for
11:    for all  $w_i$  in  $\Delta$  do
12:      if  $w_i \in D_j$  then
13:         $L_i \cup \{lbl(D_j), R\}$ 
14:      end if
15:    end for
16:  end for
17:  Shuffle and Add all  $(w_i, L_i)$  to  $PI$ 
18:  return  $PI$ 
19: end procedure

```

Algorithm 5 GenLinkedList

```

1: procedure GenLinkedList( $sk, \mathcal{I}_c, PI$ )
2:    $\mathbb{L} = \{\}$ 
3:   for all  $e \in PI$  do
4:      $w_i = e.w_i$ 
5:      $L = e.L_i$ 
6:      $len = 0$ 
7:      $cnt = \mathcal{I}_c[i][0]$ 
8:     for all  $lbl_{doc}$  &  $type \in L$  do
9:        $id_n = G_{id}(k_G, w_i || cnt)$ 
10:      if  $len == 0$  then
11:         $k_h = G_n(k_G, id_n)$ 
12:         $\mathbb{L}_s = \{\}$ 
13:      end if
14:      AddNode( $\mathbb{L}_s, k_{SE}, k_G, id_n, type$ )
15:       $len ++$ 
16:       $cnt ++$ 
17:    end for
18:     $\mathbb{L} \cup \mathbb{L}_s$ 
19:     $\mathcal{I}_c[i][0] = cnt$ 
20:     $\mathcal{I}_c[i][1] = len$ 
21:  end for
22:  return  $\mathbb{L}$ 
23: end procedure

```

We employ an ephemeral key to encrypt the private data in each node. This key will be generated using the receptive function, $k_h = G_n(k_G, id_n)$, which is shown in line 10-13 of Algorithm 5. All of the data in a linked-list will be encrypted using the same ephemeral key except the type data. We employ the secret key, k_{SE} , to encrypt the type field, because the client should be the only party who can decrypt this data. For instance, assume the client intends to create a secure linked-list for w_{21} 's list, $\{D_4, D_7\}$. Assuming $cnt = 0$, we generate the node IDs, $id_{10} = G_{id}(k_G, w_{21} || 0)$.

$id_{11} = G_{id}(k_G, w_{21} || 1)$. Afterward, we create an ephemeral key $k_h = G_n(k_G, id_{10})$ to encrypt the nodes.

Remark that for the whole linked-list, we only generate one ephemeral key (see line 10). Lastly, using AddNode, we create and encrypt the required nodes and add them to the corresponding linked-list. We demonstrate how we create a linked-list from an inverted plain-index in Algorithm 5. Once all of the nodes and required linked-list are created, the user sends them to the cloud server to be stored on the server index, \mathcal{I}_s . Remark that the index entries are encrypted and sent in an arbitrary order and the server cannot link them.

- Encryption. Using the secret key k_{SE} , the client encrypts the entire corpus (all of the files), transfers them to the cloud server including the file IDs.

Algorithm 6 $((\mathcal{I}'_c, (\mathcal{I}'_s, C')) \leftarrow \text{Update}((sk, \mathcal{I}_c, add, in), (\mathcal{I}_s, C))$ **Client**

```

1:  $\Delta_D = \text{ExtractKeywords}(D_{n+1})$ 
2:  $\mathbb{L} = \{\}$ 
3: for all  $w_i \in \Delta_D$  do
4:    $cnt = \mathcal{I}_c[i][0]$ 
5:    $len = \mathcal{I}_c[i][1]$ 
6:    $id_h = G_{id}(k_G, w_i || cnt)$ 
7:    $id_n = G_{id}(k_G, w_i || cnt - len)$ 
8:    $k_h = G_n(k_G, id_n)$ 
9:    $\mathbb{L} \cup (id_h, k_h)$ 
10:   $\mathcal{I}_c[i][0] ++$ 
11:   $\mathcal{I}_c[i][1] ++$ 
12: end for
13:  $C_{n+1} = \text{Enc}(k_{SE}, D_{n+1})$ 
14: Send  $\mathbb{L}, C_{n+1}$  to server

```

Server

```

15:  $C \cup C_{n+1}$ 
16: Update  $\mathcal{I}_s$  using  $\mathbb{L}$ 

```

- Search. Obfuscating and hiding the access and search pattern is our primary goal. To achieve this objective, we append a bounded number of sub-queries, $\mathbf{q}_i = \{q_{i1}, q_{i2}, \dots, q_{ik}\}$, to each query \mathbf{q}_i . Every sub-query consists of two tuples, (k_h, id_h) . k_h is an ephemeral key that decrypts a linked-list starting with id_h (linked-list's header). Employing \mathcal{I}_c and k_G enable the user to re-generate k_h and id_h which are required for generating the query. In addition, the noise sub-queries will be added to boost the security and privacy of the outsourced data and obfuscate the access and search pattern.

To flatten the APV and amplify the privacy of the outsourced files, we use a biased random generator function called GenRandom(). This function chooses high accessed documents with lower probability, so the less accessed files has more opportunity to be selected which impact directly the pace of flattening the APV. Note that we assign s_δ number of labels to each file. Hence, along with APV, GenRandom() inputs s_δ to randomly choose one of the available labels for the file of interest. ν holds the number of fake/noise queries that can be determined randomly. Before

sending the query to the cloud server, we insert the sub-queries in arbitrary positions in query \mathbf{q}_i .

The query generation process for a keyword w is demonstrated in Algorithm 7. Remark that the node ID (of the linked-list header) and its respective ephemeral key are regenerated in line 15. Since the users may possess various devices with different level of computation power and resources, a number of parameters including v and s_δ can be set by the client.

Once a query, \mathbf{q}_i , is received, the cloud server runs each sub-query $q_{ij} = (id_h, k_h)$ by finding id_h (header of the requested linked-list) in \mathcal{I}_s . Employing the k_h , the server then decrypts all of the nodes (except the type field) and discards all of the used index entries. All of the extracted document IDs and their respective type fields will be added to the result set. Note that the server can store the used index entries, however, it is pointless because they are already leaked and enclose no new meta-data. Lastly, the server returns a result set consists of $(R(q_{i1}), \dots, R(q_{ik}), bag)$ where $R(q_{ij})$, $1 < j \leq k$, is the q_{ij} 's resultant file IDs; and the bag is \mathbf{q}_i 's resultant encrypted files.

Once the result set of \mathbf{q}_i is received, the client who is aware of the location of the noise and real sub-queries, decrypts and separates the real results from the bag . Next, the GenLinkedList algorithm will be called to generate new entries for queried keywords in \mathbf{q}_i .

The forward privacy of our approach is guaranteed by using non-deterministic search tokens and adding random noise sub-queries. The client then updates its index, \mathcal{I}_c , and creates new node IDs and ephemeral keys for each linked-list. Note that, since the value of the cnt is updated, brand new keys and node IDs will be generated. To track the access frequency of each file, the client then updates the APV. lastly, the new index entries will be sent to the cloud server to be stored on the server index, \mathcal{I}_s .

Albeit the cloud server is aware of the relation between the last query and new entries, it cannot determine the noise keywords from the real search keyword. In addition, the node IDs and their respective keys are one-time use and vary after each search. Furthermore, there exist fake\noise nodes among the actual nodes in every linked-list. As a result, it is impossible for the server to realize the actual search and access pattern. All of these specifications in our approach guarantee the forward privacy requirement and preserving the access and search pattern. The search process is described in detail in Algorithm 7.

- Update. The update algorithm consists of two functions, del and add , as follows:

- add . In add algorithm, we first extract the keywords from the new file. The algorithm then generates a node for each keyword and adds them to the respective linked-list. Next, we encrypt the the new file using the user's secret key and transfer it to the cloud server. On the other side, the server updates the \mathcal{I}_s , once the the Update request is received. The add function is described in detail in Algorithm 6.

- del . The user creates a Update request and sets the operation to del and includes the file ID in the request to delete a file, D_k . Upon receiving the del inquiry, the cloud server deletes the respective file from its storage. Nevertheless, the corresponding

index entries cannot be removed because the server does not possess the keys.

Algorithm 7 $((\mathcal{I}'_c, D_w), (\mathcal{I}'_s)) \leftarrow \text{Search}((sk, \mathcal{I}_c, w, v, APV), (\mathcal{I}_s, C))$

Client

- 1: $counter = 0$;
- 2: $\Delta_q = \{w\}$
- 3: **while** $counter < v$ **do**
- 4: $lbl = \text{GenRandom}(APV, s_\delta)$
- 5: $i = G_w(k_G, lbl)$
- 6: **if** $w_i \notin \Delta_q$ **then**
- 7: $\Delta_q \cup w_i$
- 8: $counter++$
- 9: **end if**
- 10: **end while**
- 11: $\mathbf{q} = \{\}$
- 12: **for all** w_i **in** Δ_q **do**
- 13: $cnt = \mathcal{I}_c[i][0]$
- 14: $len = \mathcal{I}_c[i][1]$
- 15: $id_n = G_{id}(k_G, w_i || cnt - len)$
- 16: $id_h = G_{id}(k_G, w_i || cnt)$
- 17: $k_h = G_n(k_G, id_n)$
- 18: $\mathbf{q} \cup (id_h, k_h)$
- 19: **end for**
- 20: **Shuffle**(\mathbf{q})
- 21: Send \mathbf{q} to server

Server

- 22: $bag = \{\}$
- 23: **for all** q_i **in** \mathbf{q} **do**
- 24: Find respective $node$ with id_h
- 25: **while** $node \neq null$ **do**
- 26: Decrypt the $node$ using k_h in q_i
- 27: Add lbl and $type$ to $R(q_i)$
- 28: Find $next$ $node$
- 29: **end while**
- 30: Add files corresponds to $R(q_i)$ to bag
- 31: **end for**
- 32: Send $(R(q_1), \dots, R(q_k), bag)$ to client

Client

- 33: Decrypt results R
- 34: $PI = \{\}$
- 35: update APV based on the results
- 36: **for all** w_i **in** Δ_q **do**
- 37: $L =$ All doc-ids contain w_i in R
- 38: $PI \cup (w_i, L)$
- 39: **end for**
- 40: $\mathbb{L} = \text{GenLinkedList}(sk, \mathcal{I}_c, PI)$
- 41: Send \mathbb{L} to server
- 42: Delete noise results
- 43: Consume real results

Server

- 44: Update \mathcal{I}_s using \mathbb{L}

However, the index entries will be removed over time and after receiving a number of queries. The server simply removes the nodes that are pointing to a deleted file. To incorporate this

feature in Algorithm 7, the server first investigate the availability of the a file extracted from a node (line 28). The server removes them from the result set, if they do not exist.

6 Experimental results and complexity analysis

To assess the efficiency of our approach, we study and compare the complexity of the state-of-the-art methods [7], [6], [5] with our approach. Lastly, we finish this section by demonstrating the experimental results that are obtained using real world datasets.

6.1 Analyzing the complexity of our proposed algorithm

Required storage space for client and server. We first show that the amount of data that the server and especially the client should store is reasonable and manageable. Recall that user must store a dictionary, \mathcal{I}_c , on her side which holds the number of nodes (a counter) and the length of each linked-list for each keyword. Hence, the client index look likes a table with two column and m rows where m is the number of keywords. Hence, \mathcal{I}_c is an $O(m \times 2) \approx O(m)$ dictionary. Assume the user's dataset consists of 1M keywords. Moreover, suppose each integer requires 4 bytes on the memory and each keyword has an average 10 bytes. In this scenario, the user needs to store a 18 MB dictionary on her side ($1M \times (10 + 4 + 4) \approx 18MB$). Considering resource-constrained devices such as cellphones which have limited memory space and constrained computations, 18 MB is rational, cost-efficient, and manageable. As an alternative, by using the method in [7] also proposed, it is feasible to outsource the user index. In comparison to other work, in [7] the author needs $O(m + n)$, in [5] the author requires $O(\sqrt{N})$, and in [6] the author occupies $O(m)$, where n is the number of documents and N is the number of $(keyword, doc\ id)$ tuples. Regarding the size of the server index, our method needs $O(N + k)$, in which k is the number of fake/noise nodes. All state-of-the-art methods that we mentioned above require a space with size of $O(N)$ to store the server index.

Supporting parallelism by design. Beside our approach, this requirement is also fulfilled in [7] among the Dynamic SSE schemes that support forward privacy. Since in our approach the node IDs are generated by a pseudo-random function and the server index entries are independent, it is possible to distribute the sub-queries among the processors to expedite the update and search process, and achieve parallelism. The complexity of our search method is $O(d + k_d)/p$ and our update (add) system-cost is $O(r/p)$, where p is the number of cores/CPU's, d holds the number of a files containing a keyword, k_d shows the number of fake nodes in a keyword list, and r holds the number of keywords in a file. The best-case scenario happens when the number of sub-queries are equal to the number of available cores/CPU's. As a result, all sub-queries will be executed concurrently. The search cost in [7] is $O(d + n_d)/p$ and the add/update cost is $O(r/p)$, where n_d shows the number of times that a keyword has been affected by file deletions since last search. Table 1 shows our complexity analysis.

Table 1: Complexity Analysis of Related Work and Our Approach

Approach	\mathcal{I}_c	\mathcal{I}_s	Parallelism	Search	Update
Stefanov[5]	$O(\sqrt{N})$	$O(N)$	–	$O(d)$	$O(r)$
Bost[6]	$O(m)$	$O(N)$	–	$O(d)$	$O(r)$
Etemad[7]	$O(m + n)$	$O(N)$	✓	$O(d + n_d)/p$	$O(r/p)$
Ours	$O(m)$	$O(N + k)$	✓	$O(d + k_d)/p$	$O(r/p)$

6.2 Experimental results

We implemented a prototype and conducted a thorough and comprehensive evaluation to study our approach using real-life datasets. We employed Java (JDK 1.8) as the programming language and Crypto packages for the encryption process. The server and client connect and communicate through a TCP connection. Moreover, Windows machines were used for both server and client. Each machine came with 8GBs RAM and a Corei7 CPU at 3.6 GHz. To assess our scheme, we used the real-world Enron email dataset [31]. We ran each experiment ten times and the output is the average of all trials. The variance of the 10 trials were very low to be notable. We implemented the search\query algorithm twice, once using a parallel algorithm (four cores) and another time in a sequential manner. We call the former *multi-threaded* and the latter *single-threaded*. The results shows an admissible and reasonable overhead on the system that even a user with a resource-constrained device can benefit from our approach.

Table 2: # of server index entries

#Docs	#server entries
10000	829799
20000	1571676
30000	2568438
40000	3548027
50000	4404160
60000	5341524
70000	6194452

Setup time. We first started by studding the setup time per various number of files. As we discussed in Section 5.2, the setup phase includes several steps including the encryption process, creating the plain index, and the encrypted linked-lists. Our results indicate that a dataset with 20K files requires less than minute ($\approx 59\ sec$), while the same experiment, setup process, for a corpus with 50K needs less than seven minutes to be finished (see Figure 2). Remark that, this process only happens at the beginning of our approach, so it is a one-time process. In addition, we investigated the number of server index entries. Our study shows that around 4.4×10^6 entries were generated for 50K files, and 1.57×10^6 for 20K documents (see Table 2).

Query generation process time. To search for a keyword, the user needs to create a query. Each query consists of numerable search tokens\sub-queries. Every search token includes the header ID of a linked-list and its receptive key. To study the impact of number of fake/noise sub-queries on the query generation process, we queried the same keyword several times but with various number of fake keywords. The results demonstrate that the system requires less

than 1.5 milliseconds to generate a query which contains 50 fake keywords. Remark that we used 50000 files for this experiment.

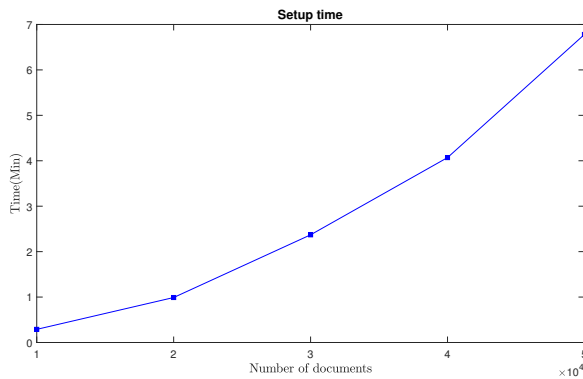


Figure 2: Client setup time

Search time. Once the server receives a search request, it will look for all files that match the search token. In this experiment we aimed to measure the amount of the time that each query requires. However, the size of the result-set (number of resultant files for a specific query) is a crucial factor in this experiment. Hence, we created ten keywords and injected them into our corpus with frequencies from 100 to 1000. These new injected keywords will enable us to investigate the effects of the number of resultant documents on the search time. Our results from the single threaded algorithm show that it takes around two seconds from the server to execute a query with 1000 resultant files. Moreover, multi-threaded algorithm requires considerably less time (around 45 percent) to answer the same query. Note that, the noise was set to three in both experiment settings.

Update (Add/Delete) requests. The user may request for an update on the corpus that can be a delete or an add request. Creating a delete query is a very low-cost operation, and takes less than 1ms in our approach. To remove a file, the user should sets the operation mode to *del* and embeds the file ID in the query. To add a new file, we first remove the stop words and extracts the main keywords. We then creates the index entries, plain index, encrypted linked-lists, and encrypt the file. Lastly, we transfer it to the cloud server. Once the cloud server receives the add query, it adds the encrypted file to the corpus and store the index entries. Note that because all of the index entries are encrypted with a new ephemeral key, the cloud server cannot determine the relation between current files in the corpus and the new file. To run our experiment, a file with 155 words and 68 keywords (excluding the stop words) from Enron dataset was selected. The operation lasts around 1.8 ms.

Obfuscating the Access Pattern. The most important goal in our approach is preserving the search and access pattern. To achieve this goal, we injected noise nodes among each linked-list's nodes, and also added noise sub queries to the main query. With this strategy, the access pattern vector (APV) become flattened and obfuscated. To measure how flattened/uniform the APV become before and after applying our approach, we used the Shannon entropy. Due to not having access to the real-life search requests, we randomly selected the queries from the keyword-set. We set the noise to three and issued 1000 queries. To calculate the entropy

improvement we employed $((e_{our} - e_{org})/e_{org}) \times 100$, where e_{org} and e_{our} are the Shannon entropy of the calculated APV before and after applying our approach. Figure ?? demonstrates our results in detail. To illustrate, applying our approach on a corpus with 30K files flattens the APV more than two times. This means our approach has made the access pattern more secure and private more than twice.

For example, exploiting our approach on a corpus of 30K documents flattens the access pattern vector more than two times. That is, the access pattern is two times more private than before applying our scheme.

7 Security proof

We defined the DSSE Security in Definition 7 and designed our dynamic SSE scheme in Section 5.2. Here, we prove that our scheme is secure using the standard simulator model.

Theorem 1 *Let $SE = (\text{Gen}, \text{Enc}, \text{Dec})$ be a CPA- secure symmetric encryption scheme, and G_n, G_{id} , and G_w be three pseudo-random functions, our DSSE scheme in Section 5.2 is secure under Definition 7.*

Proof 1 *We demonstrate how the ideal world is indistinguishable from the real world by any probabilistic polynomial time (PPT) distinguisher to prove that our dynamic and forward private SSE scheme is secure. We illustrate and explain a PPT simulator \mathcal{S} that imitate the user actions using the provided leakage functions that are defined and provided in Section 3.5. In other words, we explain how a simulator, \mathcal{S} , can adaptively mimic the user behavior including generating the encrypted indexes, queries, and documents:*

Setup. *In the first step, the simulator \mathcal{S} generates the encrypted document set, C , simulates N index entries, and creates a secret key, k_{SE} . To generate the simulated data, \mathcal{S} employs leakage function $\mathcal{L}^{\text{Setup}}(D) = \{N, n, (id(D_i), |D_i|)_{1 \leq i \leq n}\}$. Note that all data including the index entries, \mathcal{I}_s , and generated files, C , are encrypted with the secret key that was generated earlier. This means, the simulator \mathcal{S} does not require to have access to the contents of the files, and as a result, it encrypts strings of size $|D_i|$ containing all zeros to create the encrypted files. Note that no probabilistic polynomial time distinguisher (attacker) can detect and discern this behavior due to the CPA security of the applied symmetric encryption scheme. Moreover, the simulator requires to generate and keep two dictionaries, keyDict and Δ_s , to answer the Update and Search queries. Δ_s simply keeps track of the simulated keywords. For each linked-list, KeyDict dictionary stores a key, keyword, and the first node identifier of the respective linked-list. The simulator then generates the keywords and arbitrary values for linked-lists which are selected from a keyword distribution based on the range of the encryption scheme. To facilitate generating the search and update tokens, the simulator, \mathcal{S} , updates the Δ_s and keyDict dictionaries adaptively. We explained the setup phase in Algorithm 8.*

Add. *To add a new file, the simulator \mathcal{S} uses the update leakage function, $\mathcal{L}^{\text{Update}}(in, D_i, \text{op}) = \{id_{in}(D_i), |w_{in}|, |D_i|, \text{op}\}$, and employs the same keyword distribution and Δ_s . First, the simulator randomly selects $|w_{in}|$ keywords to be assigned to the new document. It then*

generates an encrypted file C_i and the respective linked-lists. Lastly, S updates the dictionaries respectively for future references. This process, add simulation, is shown in Algorithm 9. Remark that to follow our scheme's architecture, beside using a new key, every keyword is appended as a new linked-list to the cloud index. As a result, it is impossible for the server to link the newly added file to the previous search tokens even if the simulator generate a query that has the new file among the results.

Algorithm 8 Simulator's setup phase

Simulator

- 1: $k_{SE} \leftarrow \text{SE.Gen}(1^\lambda)$
- 2: Simulate C as $\{C_i \leftarrow \text{SE.Enc}(k_{SE}, \{0\}^{|D_i|})_{1 \leq i \leq n}\}$
- 3: Create *keyDict* dictionary.
- 4: Create keyword dictionary Δ_s
- 5: $\mathbb{L} = \{\}$
- 6: $node_cnt = 0$
- 7: $word_cnt = 0$
- 8: **while** $node_cnt \neq N$ **do**
- 9: $word_flag = 1$
- 10: $\Delta_s \cup w_{word_cnt}$
- 11: **while** $word_flag$ & $node_cnt \neq N$ **do** **▷ Randomly**
generates a linked-list
- 12: $list_flag = 1$
- 13: $L = \{\}$
- 14: $k_G \leftarrow \{0, 1\}^\lambda$ **▷ k_G is used to encrypt the current**
linked-list
- 15: $id_{node} \leftarrow \{0, 1\}^l$
- 16: Add $(w_{word_cnt}, id_{node}, k_G)$ into *keyDict*
- 17: **while** $list_flag$ & $node_cnt \neq N$ **do** **▷ Adds new nodes**
to L until flag becomes false
- 18: $id_{doc} \leftarrow \{id(D_i) | id(D_i) \notin L, 1 \leq i \leq n\}$
- 19: AddNode($k_G, L, id_{node}, id_{doc}$)
- 20: $node_cnt++$
- 21: $list_flag \leftarrow \{0, 1\}$
- 22: **if** $list_flag$ **then**
- 23: $id_{node} \leftarrow \{0, 1\}^l$
- 24: **end if**
- 25: **end while**
- 26: $\mathbb{L} \cup L$
- 27: $word_flag \leftarrow \{0, 1\}$
- 28: **end while**
- 29: $word_cnt++$
- 30: **end while**
- 31: Send \mathbb{L} to server

Server

- 32: Generate \mathcal{I}_s using \mathbb{L}

Search. $\mathcal{L}^{\text{Search}}$ is the leakage function that the simulator S uses to imitate the search function. This information provides enough data for the simulator to randomly selects a required number of keywords from Δ_s . In the next step, the sub-queries will be created using *keyDict* dictionary. This means, the simulator should look in the *keyDict* to find the key and node IDs for each keyword that is being searched. Once the simulator receives the results, it generates new index entries for the queried keywords and updated the

respective dictionary. We explained every step in detail in Algorithm 10. Since the queries\search tokens are non-deterministic and ephemeral, it is not feasible to unfold the search pattern using the search tokens. Moreover, recall that each sub-query can be real or fake\noise (known only to the user\simulator) which makes more difficult for the attacker to ascertain the search pattern.

Algorithm 9 Add simulation

Simulator

- 1: Simulate new file as $\{C_i \leftarrow \text{SE.Enc}(k_{SE}, \{0\}^{|D_i|})\}$
- 2: $\mathbb{L} = \{\}$
- 3: $\Delta_{tmp} = \{\}$
- 4: **for** $i = 1$ to $i < |w_{in}|$ **do**
- 5: $w \leftarrow \{w | w \in \Delta_s, w \notin \Delta_{tmp}\}$
- 6: $\Delta_{tmp} \cup w$
- 7: $L = \{\}$
- 8: $k_G \leftarrow \{0, 1\}^\lambda$
- 9: $id_{node} \leftarrow \{0, 1\}^l$
- 10: Add (w, id_{node}, k_G) into *keyDict*
- 11: AddNode($k_G, L, id_{node}, id_{doc}$)
- 12: $\mathbb{L} \cup L$
- 13: **end for**
- 14: Send \mathbb{L} to server

Server

- 15: Update \mathcal{I}_s using \mathbb{L}

Algorithm 10 Search simulation

Simulator

- 1: Generate a random value k which shows the number of keywords in the current search
- 2: $\Delta_{tmp} = \{\}$
- 3: $\mathbf{q} = \{\}$
- 4: **for** $i = 1$ to $i \leq k$ **do**
- 5: $w \leftarrow \{w | w \in \Delta_s, w \notin \Delta_{tmp}\}$
- 6: $\Delta_{tmp} \cup w$
- 7: Find w entries in *keyDict* and add them to \mathbf{q}
- 8: **end for**
- 9: Shuffle(\mathbf{q})
- 10: Send \mathbf{q} to server

Server

- 11: Perform \mathbf{q} and return the result $R = (R(q_1), \dots, R(q_k), bag)$

Simulator

- 12: Generate new \mathcal{I}_s entries based resultant *bag* from the server
- 13: Update *keyDict* respectively
- 14: Send new entries to the server

Server

- 15: Update \mathcal{I}_s according to the new entries

Hence, we programmed a simulator that mimics our approach's operations with the defined leakage functions in consideration. Remark that all simulated operations in Algorithm 8, 9, 10 are executing in polynomial time where a polynomial number of queries exists. Thus, the cloud\attacker is unable to discern the output generated by a real user from a simulator's output unless with a $neg(\lambda)$ amount or

it shatters the employed pseudo random functions or the encryption scheme.

8 Conclusions

In this paper, first, we demonstrated that DSSE schemes with forward privacy are vulnerable to leakage-abuse attacks. Moreover, we introduced two new attacks to demonstrate the vulnerability of the forward-private approaches. All SSE schemes, including approaches with forward privacy, allow a defined level of information leakage (e.g., access/search pattern) to acquire more efficiency. In our introduced attacks, we showed by reverse analyzing the access pattern, it is feasible to recover the search pattern accurately. The recovered data can be used by traditional attacks to reveal the queries, search tokens, and as a result the documents in approaches with forward privacy. Our research demonstrates that the former attacks on traditional SSE schemes are adequate to methods that follows forward privacy principals.

We then addressed this problem by constructing a new Dynamic SSE approach that support update, search, and parallelization. Our method also obfuscates the search and access pattern. In our approach, we first create an inverted-index that maps each keyword to the documents IDs containing the respective keyword. We inject fake documents' IDs in the result-set of each keyword to hide the access pattern. Only the user can discern the fake IDS from real ones. Furthermore, each search request consists of a number of sub queries where all except one are noise which is only known to the user.

Last, using a standard simulation model, we provided the security proof of our approach. Moreover, we conducted a through performance analysis on the implemented prototype that demonstrates the efficiency and low system-cost of our proposed method. As a future work, we plan to upgrade our scheme to support semi-honest cloud servers.

References

- [1] D. X. Song, D. Wagner, A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on, 44–55, IEEE, 2000.
- [2] E.-J. Goh, et al., "Secure indexes," IACR Cryptology ePrint Archive, **2003**, 216, 2003.
- [3] Y.-C. Chang, M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in International Conference on Applied Cryptography and Network Security, 442–455, Springer, 2005.
- [4] N. Cao, C. Wang, M. Li, K. Ren, W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," IEEE Transactions on parallel and distributed systems, **25**(1), 222–233, 2014, doi:10.1109/TPDS.2013.45.
- [5] E. Stefanov, C. Papamanthou, E. Shi, "Practical Dynamic Searchable Encryption with Small Leakage," in NDSS, volume 71, 72–75, 2014.
- [6] R. Bost, "Forward Secure Searchable Encryption," in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 1143–1154, ACM, 2016, doi:10.1145/2976749.2978303.
- [7] M. Etemad, A. K p c , C. Papamanthou, D. Evans, "Efficient dynamic searchable encryption with forward privacy," Proceedings on Privacy Enhancing Technologies, **2018**(1), 5–20, 2018, doi:10.48550/ARXIV.1710.00208.
- [8] X. Liu, G. Yang, Y. Mu, R. Deng, "Multi-user verifiable searchable symmetric encryption for cloud storage," IEEE Transactions on Dependable and Secure Computing, 2018, doi:10.1109/TDSC.2018.2876831.
- [9] K. Salmani, K. Barker, "Leakless privacy-preserving multi-keyword ranked search over encrypted cloud data," Journal of Surveillance, Security and Safety, 2020, doi:10.20517/jsss.2020.16.
- [10] K. Salmani, K. Barker, "Don't Fool Yourself with Forward Privacy, Your Queries STILL Belong to Us!" in Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy, CODASPY '21, 131–142, Association for Computing Machinery, New York, NY, USA, 2021, doi:10.1145/3422337.3447838.
- [11] K. Salmani, K. Barker, "Dynamic Searchable Symmetric Encryption with Full Forward Privacy," in 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP), 985–995, 2020, doi:10.1109/ICSIP49896.2020.9339338.
- [12] C. Liu, L. Zhu, M. Wang, Y.-A. Tan, "Search pattern leakage in searchable encryption: Attacks and new construction," Information Sciences, **265**, 176–188, 2014.
- [13] D. Cash, P. Grubbs, J. Perry, T. Ristenpart, "Leakage-abuse attacks against searchable encryption," in Proceedings of the 22nd ACM SIGSAC conference on computer and communications security, 668–679, ACM, 2015, doi:10.1145/2810103.2813700.
- [14] Y. Zhang, J. Katz, C. Papamanthou, "All Your Queries Are Belong to Us: The Power of File-Injection Attacks on Searchable Encryption," in 25th USENIX Security Symposium (USENIX Security 16), 707–720, USENIX Association, Austin, TX, 2016.
- [15] O. Goldreich, R. Ostrovsky, "Software protection and simulation on oblivious RAMs," Journal of the ACM (JACM), **43**(3), 431–473, 1996.
- [16] M. Naveed, "The Fallacy of Composition of Oblivious RAM and Searchable Encryption," IACR Cryptology ePrint Archive, **2015**, 668, 2015.
- [17] R. Canetti, U. Feige, O. Goldreich, M. Naor, "Adaptively Secure Multi-party Computation," in Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing, STOC '96, 639–648, ACM, New York, NY, USA, 1996, doi:10.1145/237814.238015.
- [18] X. Song, C. Dong, D. Yuan, Q. Xu, M. Zhao, "Forward private searchable symmetric encryption with optimized I/O efficiency," IEEE Transactions on Dependable and Secure Computing, 2018, doi:10.1109/TDSC.2018.2822294.
- [19] R. Curtmola, J. Garay, S. Kamara, R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," Journal of Computer Security, **19**(5), 895–934, 2011.
- [20] D. Boneh, G. Di Crescenzo, R. Ostrovsky, G. Persiano, "Public key encryption with keyword search," in International conference on the theory and applications of cryptographic techniques, 506–522, Springer, 2004.
- [21] M. Bellare, A. Boldyreva, A. O'Neill, "Deterministic and efficiently searchable encryption," in Annual International Cryptology Conference, 535–552, Springer, 2007.
- [22] N. Attrapadung, B. Libert, "Functional encryption for inner product: Achieving constant-size ciphertexts with adaptive security or support for negation," in International Workshop on Public Key Cryptography, 384–402, Springer, 2010.
- [23] A. Boldyreva, N. Chenette, Y. Lee, A. O'Neill, "Order-preserving symmetric encryption," in Annual International Conference on the Theory and Applications of Cryptographic Techniques, 224–241, Springer, 2009.
- [24] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in INFOCOM, 2010 Proceedings IEEE, 1–5, IEEE, 2010.
- [25] M. Kuzu, M. S. Islam, M. Kantarcioglu, "Efficient similarity search over encrypted data," in Data Engineering (ICDE), 2012 IEEE 28th International Conference on, 1156–1167, IEEE, 2012, doi:10.1109/ICDE.2012.23.

- [26] Z. Guo, H. Zhang, C. Sun, Q. Wen, W. Li, "Secure multi-keyword ranked search over encrypted cloud data for multiple data owners," *Journal of Systems and Software*, **137**, 380–395, 2018, doi:<https://doi.org/10.1016/j.jss.2017.12.008>.
- [27] S. K. Kermanshahi, J. K. Liu, R. Steinfeld, S. Nepal, "Generic Multi-keyword Ranked Search on Encrypted Cloud Data," in *European Symposium on Research in Computer Security*, 322–343, Springer, 2019.
- [28] S. Kamara, C. Papamanthou, T. Roeder, "Dynamic Searchable Symmetric Encryption," in *Proceedings of the 2012 ACM Conference on Computer and Communications Security, CCS '12*, 965–976, ACM, New York, NY, USA, 2012, doi:[10.1145/2382196.2382298](https://doi.org/10.1145/2382196.2382298).
- [29] M. Naveed, M. Prabhakaran, C. A. Gunter, "Dynamic Searchable Encryption via Blind Storage," in *2014 IEEE Symposium on Security and Privacy*, 639–654, 2014, doi:[10.1109/SP.2014.47](https://doi.org/10.1109/SP.2014.47).
- [30] J. Ning, J. Xu, K. Liang, F. Zhang, E.-C. Chang, "Passive attacks against searchable encryption," *IEEE Transactions on Information Forensics and Security*, **14**(3), 789–802, 2018, doi:[10.1109/TIFS.2018.2866321](https://doi.org/10.1109/TIFS.2018.2866321).
- [31] "Gutenberg Publication," <https://www.cs.cmu.edu/~enron/>, accessed: 2019-11-08.
- [32] M. S. Islam, M. Kuzu, M. Kantarcioglu, "Access Pattern disclosure on Searchable Encryption: Ramification, Attack and Mitigation." in *Ndss*, volume 20, 12, Citeseer, 2012.

Digital Competencies of Saudi University Graduates Towards Digital Society: The Case of The University of Tabuk

Inam Abousaber *

Faculty of Computers and Information Technology, Information Technology Department, The University of Tabuk, Tabuk, 71491, The Kingdom of Saudi Arabia

ARTICLE INFO

Article history:

Received: 19 February, 2022

Accepted: 18 April, 2022

Online: 22 April, 2022

Keywords:

Digital Competencies

Digital Transformation

Digital Society

Digital Citizens

ABSTRACT

This paper presents findings and proposes recommendations from an evaluation of the level of digital competencies among Saudi university graduates to ascertain their readiness for becoming digital citizens, with high confidence in using digital technologies to successfully engage in digital transformation efforts to achieve Saudi Vision 2030. The sample comprised 352 University of Tabuk students from ten different colleges (faculties) who answered an online five-point Likert-type structured questionnaire pertaining to their awareness and readiness concerning aspects of digital competence, based on previous research. The questionnaire dimensions comprised information and data literacy, digital content creation, communication and collaboration, safety, and problem-solving. The data was analyzed using SPSS statistical software package to perform, reliability test, descriptive statistics, and comparative tests. The study revealed a fair degree of digital competencies in general, with notable differences between graduates from different colleges, when comparing the five levels dimensions of digital competencies.

1. Introduction

1.1. Digital Citizens and Digital Society

The hallmark of the modern world and globalized socio-economic development is significant technological development and scientific progress, producing incredible changes in industrial practices and the everyday lives of people. Every day and ubiquitous technological tools nowadays would have been considered science fiction in the recent past, such as Smartphone technologies, virtual reality, sharing information through email to different parts of the world, etc. Technological advancement pushes society forward and makes various stakeholders increasingly dependent on digital infrastructures and technologies. Over the years, different industries have started to rely on various digital solutions and online and automated processes [1].

Governments, institutions, and business leaders all understand the importance and benefits of adopting digital changes in their respective environments to remain ahead of the competition, to ensure access to the best technology, and to make their performance more effective [2]. However, the execution of attempts to adopt and use technologies in practice often lead to

failure, representing a theory-practice gap that often frustrates the aspirations of technological investments. Consequently, researchers have explored digital transformation initiatives among both public and private sector stakeholders, to explore success and failure factors, and identify ways in which digital transformations can make workflow easier, faster, and more efficient by optimizing various processes [3].

The concept of 'digital society' is used to conceptualize and explore the interconnection between individual and community users and digital technologies deployed in various tasks [4]. Digital society refers to a society that has started to become paperless and digital, in which electronic technologies are normal and no longer seen as controversial [5]. A shift from the previous digital era has occurred as society has moved to a computational information one; indeed, society is now moving toward a new 'post-digital' world, where digital elements have been seamlessly integrated into the global economy and everyday lives [4].

Public policy responses to the Covid-19 pandemic, including social distancing and lockdown measures, galvanized latent trends in technology adoption, including the wholesale transfer of work tasks and service delivery to online platforms for certain periods (due to the lockdown of physical facilities), including offices and educational institutions. In this milieu, the concept of the 'digital citizen' has become increasingly popular to describe

*Corresponding Author: Inam Abousaber, Faculty of Computer Science & Information Technology, Information Systems Department, The University of Tabuk, Tabuk, Saudi Arabia | Email: i.abousaber@ut.edu.sa

the advancements of technology being integrated into business, work, and education, etc. [6].

In an educational context, digital citizens are normal individuals (e.g., parents and educators) who can teach in a safe online place with no ethical issues, interconnected with other teachers or students. Digital citizenship protects users, making them as safe as possible in the online world, as well as preparing students to adapt, survive, and grow in a society that relies on technology and an environment that is embedded with communications, information, and networks [6]. While Saudi Arabia is massively investing into digital transformation access in all sectors, there is a need to evaluate the level of universities' graduates to make sure they can easily integrate into the digital revelation. This paper looks at how digital transformation has occurred in Saudi Arabia and how it has affected higher education, especially among graduates at the completion of their academic programs, by testing the study Hypothesis: College of study does not significantly affect degree of digital competence.

1.2. Digital Competencies

Five digital competencies have been identified in previous literature, as explored below: information and data literacy (IDL), communication and collaboration (C&C), digital content creation (DCC), safety (SAF), and problem solving (PS) [7]

1.2.1 Information and data literacy (IDL)

IDL is the articulation of required information as well as retrieving digital data and locating information or content required for specific tasks, or whatever the user needs [8]. For example, researching immigration in China requires locating various types of official and unofficial data on immigration rates in the country, and evaluating their relative import. This type of literacy also refers to how a user judges that the data is factual and relevant to their needs [9]. Furthermore, it relates to the management, organization, and storage of data in hardware, cloud, and even paper formats, as well as how information *per se* is managed.

1.2.2 Communication and collaboration (C&C)

C&C concerns the transfer of information and interactions between people through digital technologies (e.g., using phones or on computers via Skype etc.). People communicating with one another must understand ethical, cultural, and diversity issues, and respect one another and engage in a safe manner. It allows for societies all over the world to participate in communication through private or public digital services, and can help manage their digital reputation, presence, and identity [8]. It is often used to put people out there and have their identity known. In the education sector, students close to graduation will often use digital apps such as LinkedIn to advertise their skills to companies and potential employers [9].

1.2.3 Digital Content Creation (DCC)

DCC and the editing processes are fundamental to articulating voices, such as posting stories on a short story blog or using apps such as TikTok to create whatever content a user wants to make. It relates to users understanding copyright and licensing laws, and how they are applied to any content they create, as well as

understating how the content will be used, shared, or seen, and any digital algorithms used [7].

1.2.4 Safety (SAF)

SAF This aspect pertains to protecting users and devices from harm. It protects any content users make and share with the world, and their personal data in the digital environment. It ensures that their privacy is not an issue when a user goes online [7]. It also refers to the protection and safeguarding of psychological and physical health, which is increasingly important due to cyber bullying. This also encompasses understanding that digital technologies can impact on a person's social inclusion and their social well-being. Moreover, being aware of the environmental impact that these technologies can have and how they are used also relates to SAF in the broader sense [8].

1.2.5 Problem solving (PS)

PS is for understanding any issues that could arise from a user's content or activity online [8]. It identifies how to resolve any issues in digital environments and to use digital tools to develop solutions to the problems or needs of users, such as ad blocks being created to block any ads that appear in the digital environment. It also keeps the user up to date with digital evolutions, such as updates to apps like ad blockers or new security measures, and apps that can help solve specific issues [9].

1.2.6 Digital Skills for Future Success

In [10], the author aimed to research how artificial intelligence and digital society can be used in higher education, and concluded that these digital developments were important, but that the efforts to integrate them in education are insufficient. It was noted that it is important to aid students in getting past challenges in the labor market or any workforce, as digital skills are becoming more essential for hiring agencies and employees. However, pedagogical studies have noted that teachers in educational institutions tend to resist technological integration in education, as they feel it undermines pedagogical value and human relations and feelings, making it hard to empathize with students and develop them emotionally for future preparedness and success [11]. Nevertheless, it is increasingly essential for students to have technology mixed in with their education to prepare them with prerequisite skills for a digital economy, as most careers will inevitably have various degrees of interaction with different dimensions of the digital world, thus digital skills are necessary.

Communication and information technology have been fundamentally important in career dynamics for many years, with both negative and positive influences on the work environment [12]. For example, increased internet access within a research assistant job has made it easier for people to find other papers relating to their task or find people to communicate with. This leads to employees doing their jobs accurately, using all available resources [13]. A positive pattern between earnings growth and the use of the web is apparent, indicating that behaviors and digital skills to find professionally relevant resources are rewarded by the labor market [14]. Efficiently using the internet, ensuring that any information they find is fact-checked and not false, is an important digital skill that employers look for, as false information can cause

issues within the work organization, and which can damage their reputation [13].

Trust issues in digital skills are essential for individual employees as well as organizations. Some analysts have conceptualized the division of ‘digital immigrants’ and ‘digital natives’ to refer to people who reached adulthood and began their professional lives before the popularization of mass computer and internet usage (from the 1990s onwards in developed countries). In many developing countries, this divide remains a pertinent feature of the human resource landscape. Digital immigrants spend effort and time acquiring digital skills to be able to stay at their jobs or find new jobs, as employers are looking for workers who can use technology or digital platforms accurately and efficiently. ‘Digital natives’ tend to have such skills intrinsically, although they may lack critical depth in their choices and behaviors in their general technology adoption and use. In general, however, digital skills are essential for employers when recruiting and hiring people [13].

Another skill required for future success is data intuition skills, which relate to understanding any data provided and how to apply it to solve issues within organizations, as well as the ability to develop critical thinking skills. Data visualization and communication is critical to learn, as it helps with making and supporting data-based decisions [15]. For example, the financial industry requires these skills to be able to describe why their solution works against any problems they encounter. It further develops critical thinking, which is a transferable skill for any role. Being proficient at using programming language and tools allows one to access, present or retrieve data in a neat way. Presenting data in a clear and readable format is something employees look for as it shows one can easily explain their data and present it in a way everyone can understand [15].

1.3 Digital Transformation and Saudi Vision 2030

Digital advances are intrinsically beneficial, but their application and use can pose disadvantages and disruptions. Thus governments, researchers, and industries are trying to figure out how the world will transform and change to be able to find opportunities and tackle emergent challenges. This phenomenon is known as digital transformation, which can be defined as digital usages developed to encourage creativity, innovation, and results, in significant alteration to the information or professional domains [16]. The importance of digital transformation (which is part of the Vision 2030) is being increasingly recognized. Saudi Arabia is one of the biggest technology and information markets in the Middle East and North Africa region, with public spending of USD 45 billion at the end of 2019, mainly in the education sector. During 2014–2020 the government spent around SAR 200 billion per year on X [17], with a conscious goal to promote digital technology use as per Vision 2030 [18]. The Saudi Ministry of Education implemented a digital technology system in 2020 to minimize the effect of Covid-19 on student education.

Economic diversification and private sector growth are an essential objective of Vision 2030 and the sustainable development of Saudi Arabia. The Vision seeks to drive a digital transformation and society to produce a knowledge economy, with guidelines and safety for technology users, veering

information use, processes, and policies [19]. In 2020, Saudi Arabia launched the National Transformation Program to build an institutional network with the capabilities needed to achieve the 2030 goals. This Program will help to build and develop the digital infrastructure, which is a goal for the Vision 2030. It seeks to provide a secure and powerful online platform that can be used by many people simultaneously, without the fear of the entire network failing.

Vision 2030 aims for Saudi Arabia to become a global investment powerhouse, hence it focuses on empowering graduates with digital skills that can be used in any sector or industry, to increase employability and national human resources, which can attract global investment. The goal is to also transform the public investment fund into the largest sovereign wealth fund in the world. This was spearheaded by the floating of Saudi Aramco, and is now diffused into local production, economic expansion, workforce expansion, and sending students to study abroad, etc. [20]. By 2030, Saudi Arabia wants to increase the involvement and achievements of its young workforce, particularly graduates and students. It has invested massively in domestic education and scholarships for advanced studies by young Saudi students [17]. These invitations will contribute to the country’s economy and increase the skills of the local workforce [20].

1.4 Efforts and Importance of Empowering Saudi University Graduates with Digital Skills

Studies understand that their future success entails training in digital skills, and Saudi national educational policy and Vision 2030 recognizes the impact of technology on learning and e-learning potential [21]. E-services have an increasing presence on Saudi campuses and within lessons, and it is essential to ensure that students know how to use them, and what they can be used for within work environments. 70.3% of staff members are also aware of the advantages of their students having digital skills and the importance of practicing them in the field (e.g., a work environment) [22].

While Vision 2030 has galvanized technology adoption in Saudi education, this is merely the latest stage of a long-term overhaul of the national educational paradigm, which began in 2005 when King Abdullah agreed with President G. W. Bush to increase the number of male and female Saudi students studying in the US, thereby inaugurating the King Abdullah Scholarship Program (KASP), targeted to facilitate the development of Saudi human resources [23]. This also helped alleviate the pressure of unemployment due to limited opportunities for work in the private sector at the time, thus sending students to study abroad to learn and acquire skills was doubly effective, in training the future workforce and alleviating contingent pressures [24]. This was the largest scholarship program in Saudi history, seeking to prepare future generations for future careers and a knowledge-based society [23].

The National Transformation Program continues these goals, sending students to the best universities to learn from the best educators, setting high professional and academic standards, and exchanging digital, educational, cultural, and scientific experience with different countries [25]. This helps develop the

workforce level of professionalism and skills for employability and flexibility [22]. It also seeks to reduce the gender gap within the workforce and education, particularly by empowering women with education, learning skills to be able to work in whatever field they want. 82% of students and graduates on the Program believed that studying abroad and getting foreign qualifications would result in higher paid jobs for them when they returned to Saudi Arabia [25].

Similarly, the Cooperative Training Strategy combines programs among employment agencies and universities to give students a way to put their skills in use and learn new transferable skills. It builds technical, social, and moral qualities that can be used in future jobs. The objective of the program is to explore a variety of different jobs while studying, to understand which sector suits them or interests them. It emphasizes cooperative knowledge of the working environment and experiential learning while developing skills [22].

2. Methodology

This study uses a quantitative research method with a cross-sectional survey, to obtain university graduates' opinions on the subject being tested at the stage of their graduation, the given time for the study [26]. Electronic survey was the most suitable method, as it enables quick data collection, reachability, low cost, and direct digitization of data to be transferred to data analysis software. This was also expedient and safe due to on-going Covid-19 restrictions at the time of data collection.

2.1 Participants

The survey targeted 6,000 graduates according to the university's expected number of graduates for the last three years. Graduates from ten different colleges were invited to take part; the highest response came from the College of Education and Arts (17.3%), while the lowest response was from the College of Pharmacy (4.3%). The distribution of responses by college of study is shown in Table 1. The total number of respondents was 352, which is considered a representative sample for this study.

Table 1: Participant colleges of study

College of Study	N	%
College of Education and Arts (CoEA)	61	17.3
College of Sharia and Regulations (CoSR)	44	12.5
College of Computer and Information Technology (CoCIT)	41	11.6
College of Business Administration (CoBA)	39	11.1
College of Science (CoSc)	39	11.1
College of Medicine (CoMed)	33	9.4
College of Art and Design (CoAD)	29	8.2
College of Engineering (CoEng)	29	8.2
College of Applied Medical Sciences (CoApMed)	22	6.3
College of Pharmacy (CoPharm)	15	4.3
Total	352	100

2.2 Data Collection Tool

The Digital Competence Questionnaire was adopted from tools used in previous literature [27] [28], which in turn were based on the DigComp framework [29]. It was digitized on Google Forms and was used as a data collection tool. The questionnaire included two main sections: the general information section, including

college selection, and the section covering the five analyzed dimension of digital competence (IDL, C&C, DCC, SAF, and PS).

The questionnaire comprised paragraph items answerable with a five-point Likert-type scale, ranging from 1 ("strongly disagree") to 5 ("strongly agree"). The instrument was validated by two members of staff from the CoCIT to make sure it works properly; it was found to be clear and understandable for a wide range of participants from different colleges.

2.3 Data Analysis

SPSS software package was used to perform the analysis. Three types of analysis methods were used in this study: reliability test, aiming to make sure the responses are valid before meaningful analysis; descriptive statistics, to provide general overview of on the level five aspects of digital competence by presenting the mean for each item in questionnaire; and comparative analysis, to examine differences between the ten groups' data. The assumptions reading the analysis were established before the analysis started. The means, standard deviation (SD), frequency, percentages, and degrees were calculated based on the following:

Length of period	of	=	Upper bound - lower bound	=	5-1	1.33
			Number of levels		3	

The number of period levels was thus as follows: low (1-2.33), medium (2.34-3.67), and high (3.68-5). These values were in the ranges stated by [30] and can be considered as normal distribution. Similarly, the normality of distribution for each sub-group was examined. Cronbach's alpha coefficient was used to test the stability of the study instrument, requiring a score of at least (0.6) to indicate that items measured the variables they were intended to, and that the instrument was consistent and dependable. The Cronbach's alpha coefficient (0.83) indicates the stability of the study tool [31]. One-way analysis of variance (ANOVA) was used to test the difference between the levels of digital competence between different groups (i.e., different colleges).

3. Results and Discussion

3.1 Descriptive Analysis

The scores shown in Table 2 fundamentally answer the question of the degree of digital competence among participants. It shows the mean scores of all items representing digital competence; all items were measured using a five-point Likert scale. There is a good general indication that participants had a high level of competence for nearly two-thirds of items, but all of them were marginal at the lower high. While the other third of the items scored medium.

Table 2: Digital Competence Score

Digital Competence	Mean	SD	%	Degree
When sharing my personal information online, I take precautions to protect the personal data of others (not to tag them in a photo without permission, etc.)	3.86	1.163	77.2	High
I am aware of the risks and threats in online environments	3.84	1.239	76.8	High

Digital Competence	Mean	SD	%	Degree
I take precautions about safety and privacy in online environments	3.82	1.211	76.4	High
I comply with behavioral norms (ethical rules) when interacting in online environments	3.81	1.18	76.2	High
I am aware of the effects of digital technology use on health (physical, psychological)	3.8	1.174	76	High
I protect personal data and privacy in online environments	3.79	1.162	75.8	High
I investigate from different sources whether the data, information or digital content I access is reliable	3.78	1.227	75.6	High
I identify my needs when searching for data, information, or digital content in online environments	3.77	1.206	75.4	High
I am familiar with data policies (how to use personal data) of the digital services that I am a user of (social networking, etc.)	3.76	1.186	75.2	High
I take different measures to protect my digital device and content	3.76	1.191	75.2	High
I am aware of the environmental impact of using digital technologies	3.74	1.178	74.8	High
I use digital technologies to communicate in online environments	3.72	1.113	74.4	High
I can develop content in different formats (video, visual, animation, etc.) using digital technologies	3.72	1.212	74.4	High
I develop my digital competence by following new developments	3.71	1.156	74.2	High
I share data, information or digital content using different digital technologies	3.7	1.108	74	High
I easily organize and store data, information and content in online environments	3.7	1.146	74	High
I pay attention to source and citation representations when sharing data, information or digital content	3.69	1.138	73.8	High
I access the data, information and digital content I need in online environments	3.68	1.078	73.6	High
I know what to look out for when creating a digital identity (profile) in online environments	3.68	1.204	73.6	High
I know how to deal with online threats	3.66	1.176	73.2	medium
I pay attention to copyrights and licensing when developing digital content	3.65	1.169	73	medium
I identify the causes of technical problems I encounter when using digital media and devices	3.65	1.167	73	medium
I use digital technologies to collaborate in online environments	3.65	1.147	73	medium
I am aware that I leave a digital footprint when I navigate online environments	3.62	1.118	72.4	medium
I use information search strategies to access data, information, and digital content in online environments	3.61	1.129	72.2	medium
I develop content in simple forms using digital technologies	3.61	1.12	72.2	medium
I produce digital content by making changes to ready-made content	3.6	1.143	72	medium
I solve the technical problems I encounter when using digital media and devices	3.59	1.072	71.8	medium
I identify opportunities for the development of my digital competences	3.58	1.101	71.6	medium

Digital Competence	Mean	SD	%	Degree
I use different digital technologies to create innovative solutions	3.57	1.155	71.4	medium
I critically evaluate the accuracy of the data, information or digital content I access	3.54	1.149	70.8	medium
Average	3.70	1.159	74.0	High

Table 3 shows that participants had higher digital competence scores for the SAF dimension (3.76), followed by DCC (3.72), and IDL (3.68); and they had medium scores for C&C (3.64) and PS (3.62).

Table 3: Scores for Digital Competence Items

Digital Competence	Mean	SD	%	Degree	Ranking
IDL	3.68	0.968	73.6	Medium	3
DCC	3.72	0.986	74.4	High	2
C&C	3.64	1.016	72.9	Medium	4
SAF	3.76	1.023	75.1	High	1
PS	3.62	0.998	72.4	Medium	5
Average	3.7	1.159	74	High	

The following analysis discusses each dimension separately.

Table 4 shows that participants had high scores for all but two items representing IDL (3.68-3.78), and the paragraph “I investigate from different sources whether the data, information or digital content I access is reliable” had the highest degree. The paragraphs “I use information search strategies to access data, information, and digital content in online environments” and “I critically evaluate the accuracy of the data, information or digital content I access” had medium scores (3.61, 3.54, respectively).

Table 4: Scores for IDL Items

Paragraph	Mean	SD	%	Degree
I investigate from different sources whether the data, information or digital content I access is reliable	3.78	1.227	75.6	High
I identify my needs when searching for data, information or digital content in online environments	3.77	1.206	75.3	High
I pay attention to source and citation representations when sharing data, information or digital content	3.69	1.138	73.9	High
I access the data, information and digital content I need in online environments	3.68	1.078	73.7	High
I use information search strategies to access data, information, and digital content in online environments	3.61	1.129	72.3	Medium
I critically evaluate the accuracy of the data, information or digital content I access	3.54	1.149	70.9	Medium
Average	3.68	.968	73.6	Medium

Table 5 shows that all items representing C&C got high scores (3.70-3.81), except for the medium score for “I use digital technologies to collaborate in online environments” (3.65). The paragraph “I comply with behavioral norms (ethical rules) when interacting in online environments” had the highest score.

Table 5: Scores for C&C Items

Paragraph	Mean	SD	%	Degree
I comply with behavioral norms (ethical rules) when interacting in online environments	3.81	1.180	76.1	High
I use digital technologies to communicate in online environments	3.72	1.113	74.4	High
I share data, information or digital content using different digital technologies	3.70	1.108	74.1	High
I easily organize and store data, information and content in online environments	3.70	1.146	74.1	High
I use digital technologies to collaborate in online environments	3.65	1.147	73.0	Medium
Average	3.72	.986	74.4	High

Table 6 shows that for all items representing DCC participants had medium scores (3.60-3.65), except the paragraph “I can develop content in different formats (video, visual, animation, etc.) using digital technologies,” which had a high score (3.72).

Table 6: Scores for DCC Items

Paragraph	Mean	SD	%	Degree
I can develop content in different formats (video, visual, animation, etc.) using digital technologies	3.72	1.212	74.3	High
I pay attention to copyrights and licensing when developing digital content	3.65	1.169	73.0	Medium
I develop content in simple forms using digital technologies	3.61	1.120	72.1	Medium
I produce digital content by making changes to ready-made content	3.60	1.143	72.0	Medium
Average	3.64	1.016	72.9	Medium

Table 7 shows that participants had high scores for most items representing SAF (3.68-3.86), with the highest score for the paragraph “When sharing my personal information online, I take precautions to protect the personal data of others (not to tag them in a photo without permission, etc.)” Medium scores were reported for the items “I know how to deal with online threats” (3.66), and “I am aware that I leave a digital footprint when I navigate online environments” (3.62).

Table 7: Scores for SAF Items

Paragraph	Mean	SD	%	Degree
When sharing my personal information online, I take precautions to protect the personal data of others (not to tag them in a photo without permission, etc.)	3.86	1.163	77.2	High
I am aware of the risks and threats in online environments	3.84	1.239	76.9	High
I take precautions about SAF and privacy in online environments	3.82	1.211	76.3	High
I am aware of the effects of digital technology use on health (physical, psychological)	3.80	1.174	76.0	High
I protect personal data and privacy in online environments	3.79	1.162	75.9	High
I am familiar with data policies (how to use personal data) of the digital services that I am a user of (social networking, etc.)	3.76	1.186	75.3	High

I take different measures to protect my digital device and content	3.76	1.191	75.2	High
I am aware of the environmental impact of using digital technologies	3.74	1.178	74.9	High
I know what to look out for when creating a digital identity (profile) in online environments	3.68	1.204	73.5	High
I know how to deal with online threats	3.66	1.176	73.1	Medium
I am aware that I leave a digital footprint when I navigate online environments	3.62	1.118	72.4	Medium
Average	3.76	1.023	75.1	High

Table 8 shows that participants had medium scores for all items representing PS (3.57-3.65), except the paragraph “I develop my digital competence by following new developments,” which had a higher score (3.71).

Table 8: Scores for PS Items

Paragraph	Mean	SD	%	Degree
I develop my digital competence by following new developments	3.71	1.156	74.1	High
I identify the causes of technical problems I encounter when using digital media and devices	3.65	1.167	73.0	Medium
I solve the technical problems I encounter when using digital media and devices	3.59	1.072	71.8	Medium
I identify opportunities for the development of my digital competences	3.58	1.101	71.6	Medium
I use different digital technologies to create innovative solutions	3.57	1.155	71.4	Medium
Average	3.62	.998	72.4	Medium

3.2 Comparative Analysis: Hypothesis Testing

Hypothesis: College of study does not significantly affect degree of digital competence ($\alpha \leq 0.05$).

To test the above hypothesis, we used one-way ANOVA. Table 9 shows that all (F) values were statistically significant at ($\alpha \leq 0.05$) except for C&C, thus we conclude that there is a statistically significant difference in degree of digital competence by college of study at ($\alpha \leq 0.05$) for IDL, DCC, SAF, and PS, with superior competence among respondents from CoCIT, and the worst among CoAD, as shown in Figure 1.

Table 9: One-way ANOVA

Study College	N	Mean	SD	df	F	Sig.
Information and data literacy (IDL)						
CoApMed	22	3.80	.854	9	2.375	.013*
CoAD*	29	3.29	.931			
CoBA	39	3.56	1.020			
CoCIT*	41	4.00	.911			
CoEA	61	3.42	1.089			
CoEng	29	3.82	.920			
CoMed	33	3.58	.978			
CoPharm	15	3.48	.919			
CoSc	39	3.90	.947			
CoSR	44	3.90	.771			
Total	352	3.68	.968			
Communication and collaboration						
CoApMed	22	3.57	.957	9	1.123	.345
CoAD	29	3.46	.909			
CoBA	39	3.58	1.000			

Study College	N	Mean	SD	df	F	Sig.
CoCIT	41	3.94	1.021			
CoEA	61	3.58	1.097			
CoEng	29	3.79	.871			
CoMed	33	3.67	1.016			
CoPharm	15	3.64	.914			
CoSc	39	3.90	1.064			
CoSR	44	3.93	.814			
Total	352	3.72	.986			
Digital content creation						
CoApMed	22	3.41	1.010	9	2.473	.010*
CoAD*	29	3.28	.963			
CoBA	39	3.48	.984			
CoCIT*	41	4.05	.978			
CoEA	61	3.44	1.088			
CoEng	29	3.86	.898			
CoMed	33	3.48	1.066			
CoPharm	15	3.52	.837			
CoSc	39	3.94	1.065			
CoSR	44	3.80	.903			
Total	352	3.64	1.016			
Safety						
CoApMed	22	3.66	1.004	9	2.499	.009*
CoAD	29	3.51	1.047			
CoBA	39	3.66	1.038			
CoCIT	41	4.14	.943			
CoEA	61	3.46	1.195			
CoEng	29	3.90	.963			
CoMed	33	3.56	1.080			
CoPharm	15	3.50	.891			
CoSc	39	3.99	.915			
CoSR	44	4.05	.752			
Total	352	3.76	1.023			
Problem solving						
CoApMed	22	3.47	1.066	9	3.092	.001*
CoAD	29	3.27	1.165			
CoBA	39	3.38	1.020			
CoCIT	41	4.07	.942			
CoEA	61	3.39	.987			
CoEng	29	3.81	.982			
CoMed	33	3.47	.996			
CoPharm	15	3.36	.836			
CoSc	39	3.86	.927			
CoSR	44	3.90	.790			
Total	352	3.62	.998			
At						
CoApMed	22	3.61	.908	9	2.511	.009*
CoAD*	29	3.39	.946			
CoBA	39	3.56	.974			
CoCIT*	41	4.06	.895			
CoEA	61	3.46	1.046			
CoEng	29	3.85	.872			
CoMed	33	3.56	.980			
CoPharm	15	3.50	.847			
CoSc	39	3.93	.909			
CoSR	44	3.94	.711			
Total	352	3.70	.940			

* statistically significant at ($\alpha \leq 0.05$)

4. Conclusion and Recommendations

The outcomes from this study provide some indications of the level of digital competencies among Saudi graduates as future digital citizens joining the local and global digital transformation. The findings showed some good general indications when looking at all items of all five indicators used to access participant levels of digital competencies. However, when comparing findings for each indicator, moderate competence levels are evident for three out of five studied dimensions: IDL, C&C, and PS. Differences

between graduates from different colleges are evident when comparing levels of digital competencies. Hence, it could be recommended to conduct more studies with different universities to obtain comparative findings to confirm these academic specialty-related differences in other institutions to address this study limitations. Moreover, efforts must be made to sustain and promote good areas of digital competencies, with targeted strategies and support for learners with specific needs to increase their competence in certain areas. A national level of digital competencies assessment framework could be linked with other efforts of Saudi digital transformation mission to help continuous evaluation of graduate levels and provide suggestions to accommodate enhancement to enable abilities for the use of future technologies. This would help integrate pedagogical efforts toward achievement of Saudi Vision 2030 and improve the employability of Saudi graduates.

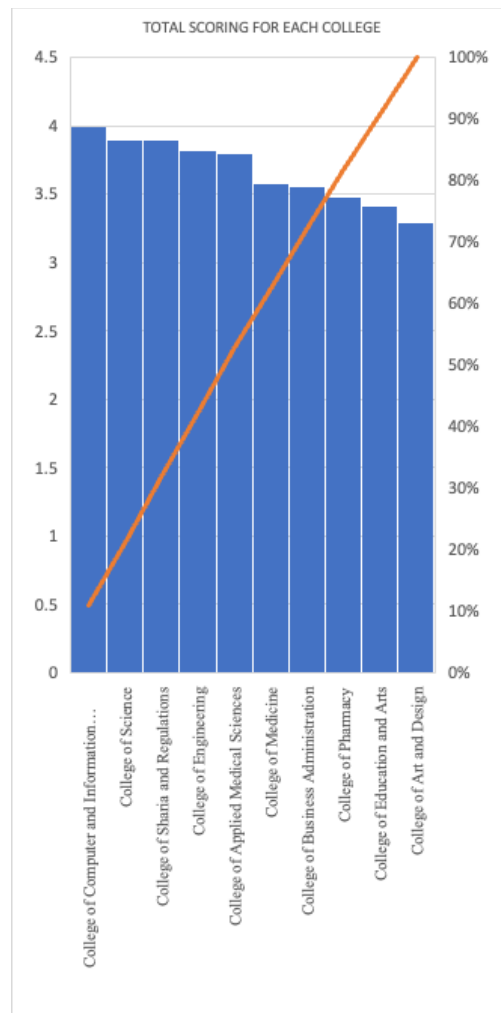


Figure 1: Total scoring for each college.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgment

The author would like to thank all students at the University of Tabuk who participated in this study.

References

- [1] A. Mohammed, R. Ali, A. Abdulkareem, B. Alharbi, "The reality of using artificial intelligence techniques in teacher preparation programs in light of the opinions of faculty members: a case study in Saudi Qassim University," *Multicultural Education Journal*, **7**(1), 5-16, 2021, doi:10.5281/zenodo.4410582
- [2] J. L. Morrison, "Organizational change for corporate sustainability: A guide for leaders and change agents of the future," *Journal of Education for Business*, **79**(2), 124-126, 2003, doi:10.1080/08832320309599100
- [3] A. Omar, A. Almaghthawi, "Towards an integrated model of data governance and integration for the implementation of digital transformation processes in the Saudi universities," *International Journal of Advanced Computer Science and Applications*, **11**(8), 588-593, 2020, doi:10.14569/IJACSA.2020.0110873
- [4] D. Berry, *The Philosophy of Software: Code and Mediation in the Digital Age*. Amsterdam: Springer, 2016.
- [5] A. Martin, Digital literacy and the 'digital society', In C. Lankshear & M. Knobel(Eds), *Digital Literacies: Concepts, Policies, and Practices*, 151-176, Peter Lang, 2008.
- [6] T. Tan, "Educating digital citizens," *Leadership*, **41**(1), 30-32, 2011, ISSN-1531-3174, Retrieved from: <https://files.eric.ed.gov/fulltext/EJ965963.pdf>
- [7] S. Carretero, R. Vuorikari, Y. Punie, *The Digital Competence Framework for Citizens*, Publications Office of the European Union, 2017.
- [8] M. Fucci, *The Digital Competence Framework 2.0* [online]. Brussels: European Commission, EU Science Hub, 2015. Available at: <https://ec.europa.eu/jrc/en/digcomp/digital-competence-framework#:~:text=%20DigComp%202.0%20identifies%20the%20key%20components%20of,Accessed 21 Jun. 2021>.
- [9] L. Ilomäki, A. Kantosalo, and M. Lakkala, *What is Digital Competence?* [online]. Brussels: European Schoolnet, 2011, Available at: https://helda.helsinki.fi/bitstream/handle/10138/154423/Ilom_ki_etal_2011_What_is_digital_competence.pdf?sequence=1 [Accessed 21 Jun. 2021].
- [10] H. E. Pence, "Artificial intelligence in higher education: New wine in old wineskins?," *Journal of Educational Technology Systems*, **48**(1), 5-13, 2019, doi.org/10.1177/0047239519865577
- [11] H. I. Haseski, "What do Turkish pre-service teachers think about artificial intelligence?," *International Journal of Computer Science Education in Schools*, **3**(2), 3-23, 2019, doi:10.21585/ijeses.v3i2.55
- [12] Y. Lipshits-Braziler, M. Tatar, I. Gati, "The Effectiveness of Strategies for Coping With Career Indecision: Young Adults' and Career Counselors' Perceptions," *Journal of career development*, **44**(5), 453-468, 2017, doi:10.1177/0894845316662705.
- [13] S. Lissitsa, S. Chachashvili-Bolotin, Y. A. Bokek-Cohen, "Digital skills and extrinsic rewards in late career," *Technology in Society*, **51**, 46-55, 2017, doi:10.1016/j.techsoc.2017.07.006
- [14] P. DiMaggio, B. Bonikowski, "Make money surfing the web? The impact of Internet use on the earnings of US workers," *American Sociological Review*, **72**(2), 227-250, 2008, doi:10.1177/000312240807300203
- [15] Y. Punie, K. Ala-Mutka, "Future learning spaces: New ways of learning and new digital skills to learn", *Nordic Journal of Digital Literacy*, **2**(4), 210-225, 2007, doi:10.18261/ISSN1891-943X-2007-04-02
- [16] C. Lankshear, M. Knobel, Introduction: Digital literacies-Concepts, policies and practices. In C. Lankshear & M. Knobel (Eds.), *Digital literacies: Concepts, policies and practices*, 1-16, Peter Lang, 2008.
- [17] H. Abdulrahim, F. Mabrouk, "COVID-19 and the digital transformation of Saudi higher education," *Asian Journal of Distance Education*, **15**(1), 291-306, 2020, ISSN 1347-9008, Retrieved from <https://files.eric.ed.gov/fulltext/EJ1289975.pdf>
- [18] A. Aldiab, H. Chowdhury, A. Kootsookos, F. Alam, "Prospect of eLearning in higher education sectors of Saudi Arabia: A review," *Energy Procedia*, **110**, 574-580, 2017, doi:10.1016/j.egypro.2017.03.187
- [19] M. Al-Ruithe, E. Benkhelifa, "Cloud data governance in-light of the Saudi Vision 2030 for digital transformation," in 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), 1436-1442, 2017, doi:10.1109/AICCSA.2017.217
- [20] B. Mitchell, A. Alfuraih, "The Kingdom of Saudi Arabia: Achieving the aspirations of the National Transformation Program 2020 and Saudi Vision 2030 through education," *Journal of Education and Development*, **2**(3), 36, 2018, doi:10.20849/jed.v2i3.526.
- [21] K. Al-Badawi, H. Ahmed, "A Study on the Impact of the Use of Modern Technology on Testing and Electronic Evaluation of the Student as a Means of Improving the Quality of University Education in Saudi Arabia," in 2014 6th Annual Conference of the Arab Organization for Quality Assurance(AROQA), 226-241, 2014, Retrieved from [https://events.aroqa.org/uploads/newsImage/file/final_proceedings_2014_\(1\).pdf](https://events.aroqa.org/uploads/newsImage/file/final_proceedings_2014_(1).pdf)
- [22] H. A. Ibeaheem, S. Elawady, W. Rigmoun, "Saudi Universities and higher education skills on Saudi Arabia," *International Journal of Higher Education Managemnet*, **4**(2), 69-82, 2018, doi:10.24052/IJHEM/V04N02/ART05
- [23] C. Taylor, W. Albasri, "The Impact of Saudi Arabia King Abdullah's Scholarship Program in the U.S.," *Open Journal of Social Sciences*, **2**(10), 109-118, 2014, doi:10.4236/jss.2014.210013
- [24] K. T. Hilal, S. Scott, N. Maadad, "The political, socio-economic and sociocultural impacts of the King Abdullah Scholarship Program (KASP) on Saudi Arabia," *International Journal of Higher Education*, **4**(1), 254-267, 2015, doi:10.5430/ijhe.v4n1p254
- [25] M. Alsqoor, "How has the King Abdullah Scholarship Program enhanced the leadership skills of Saudi female beneficiaries?," Ph.D. Thesis, Univeristy of St. Thomas, 2018.
- [26] J. R. Fraenkel, N. E. Wallen, H. H. Hyun, *How to Design and Evaluate Research in Education*, 8th ed., McGraw Hill, 2012.
- [27] A. Çebi, I. Reisoğlu, "A training activity for improving the digital competences of pre-service teachers: The views of pre-service teacher in CEIT and other disciplines," *Educational Technology Theory and Practice*, **9**(2), 539-565, 2019, doi:10.17943/etku.562663
- [28] A. Çebi and I. Reisoğlu, "Digital competence: A study from the perspective of pre-service teachers in Turkey," *Journal of New Approaches in Educational Research*, **9**(2), 294-308, 2020, doi: 10.7821/naer.2020.7.583
- [29] A. Ferrari, *Digcomp: A Framework for Developing and Understanding Digital Competence in Europe*, Scientific and Policy Report by the Joint Research Centre of the European Commission, Publications Office of the European Union, 2013, doi:10.2788/52966
- [30] B. G. Tabachnick, L. S. Fidell, *Using Multivariate Statistics*, 6th ed., Pearson Education, 2013.
- [31] U. Sekaran, R. Bougie, *Research Methods For Business: A Skill-Building Approach*, 6th ed., John Wiley & Sons, 2016.

Cloud-Based Hierarchical Consortium Blockchain Networks for Timely Publication and Efficient Retrieval of Electronic Health Records

Alvin Thamrin, Haiping Xu*, Rui Ming

Computer and Information Science Department, University of Massachusetts Dartmouth, Dartmouth, MA 02747, USA

ARTICLE INFO

Article history:

Received: 30 February, 2022

Accepted: 17 April, 2022

Online: 22 April, 2022

Keywords:

Hierarchical blockchain

Cloud Computing

Timely publication

Electronic health records

Consensus mechanism

ABSTRACT

Blockchain technology is seeing a trend of popularity and adoption in many different application areas. One such area is healthcare, as there is a need to develop a system that can reliably store and share electronic health records (EHRs) among hospital-based health facilities. In this paper, we present a cloud-based hierarchical consortium blockchain framework for storing and sharing EHRs in a scalable, secure, and reliable manner. The framework enables data sharing between local hospital blockchain networks (HBNs) through high-level blockchain networks, namely, city blockchain networks (CBNs) and a state blockchain network (SBN). To support the timely publication of EHRs in HBNs, we adopt a temporary and permanent block scheme in hospital blockchains. In addition, we develop role-based access control (RBAC) policies for data authorization and procedures for concurrent search and retrieval of EHRs across cities and states. The experimental results show that our proposed approach is feasible and supports timely publication and efficient retrieval of EHRs in cloud-based hierarchical blockchain networks.

1. Introduction

Blockchain technology was originally proposed in 2008 as a decentralized and distributed digital ledger mechanism for the peer-to-peer electronic cash system called Bitcoin [1]. A blockchain stores data in blocks that are cryptographically chained together in the form of a linked list. Thus, blocks can be used to store and record transactions in a tamper-proof and immutable manner. Unlike public blockchains, a consortium blockchain is defined as a permissioned blockchain, and access to it is usually restricted to a specific number of “permissioned” nodes [2]. In recent years, the popularity of consortium blockchain has increased due to its potential use in many different application areas, including healthcare [3], [4]. A consortium blockchain-based system can be implemented within the healthcare domain to enable and support the storage and sharing of healthcare data among health institutions or hospitals. In our earlier work, we introduced a cloud-based blockchain solution for storing and sharing electronic health records (EHRs) while enabling data accessibility, redundancy, and security on a local scale [5]. This solution allows storing big data, such as EHRs with multimedia files, in a cloud-based blockchain, while storing their metadata in a lite blockchain for efficient information retrieval. However, due

to the big data involved, the solution can only be effective when implemented on a small/local scope. This is because the growth potential of the blockchain increases dramatically with the large number of hospitals participating in the network. This can lead to a very unsustainable expansion of the blockchain in terms of size, which constitutes a major scalability issue.

In this paper, we present a cloud-based hierarchical consortium blockchain framework to address the above scalability issue. The framework consists of three layers of blockchain networks, namely hospital blockchain networks (HBNs), city blockchain networks (CBNs), and a state blockchain network (SBN). An HBN is designated as a blockchain network at the first layer and is shared by hospitals that are geographically close to each other in a local area or a city. To simplify matters, in this paper we use the term *city* to refer to a city, a local area or any form of governmental jurisdiction below the state level. A CBN is designated as a blockchain network at the second layer. Unlike an HBN, a CBN is shared by all cities located within a state as participants. Each city in a CBN is also connected to an HBN as the network regulator, which allows agents from different HBNs within the same state to communicate with each other for data sharing purposes. Finally, An SBN is designated as a blockchain network at the third layer. The SBN is shared by all states located within a country. Each state in the SBN is also connected to a CBN and acts as the network regulator of the CBN. The SBN is designed to allow agents from

*Corresponding Author: Haiping Xu, University of Massachusetts Dartmouth, Dartmouth, MA 02747, Email: hxu@umassd.edu

different CBNs across the country to communicate with each other for data sharing purposes, similar to the sharing relationship between a CBN and the HBNs connected under it. The implementation of these network layers enables all hospital peers across the country to communicate and share data with each other in a scalable, secure, and reliable manner.

Another challenging issue we face concerns the publishing of EHRs to the blockchain in a timely and space-efficient manner. Whenever data are made available, they can either be published to the blockchain immediately in the form of block records stored in a block, or they can be accumulated until the block contains a sufficient number of block records to be published. The first method excels in terms of timeliness but is inefficient in terms of space/memory usage because this method generates many blocks containing a single or very few records. On the other hand, the second method is more spatially efficient compared to the first one because fewer and denser blocks are generated; however, it has a significant drawback in terms of timeliness that may affect the effectiveness of a blockchain-based system for storing and sharing EHRs. In this paper, we present an approach that facilitates the timely and space-efficient publication of new block records using a temporary and permanent block scheme. As demonstrated in previous work [6], a new block record can be published immediately in a temporary block after being approved using a temporary block consensus mechanism. Once a predefined number of temporary blocks have been published, they can be merged into a permanent block and published to the blockchain.

This work significantly extends our previous proposed framework for healthcare data storage using hierarchical cloud-based blockchains. In our previous work [7], we focused on the structural design of the storage system and did not fully consider the scalability of HBNs with many peers and the timely publication of EHRs in HBNs. To address these issues, we now limit the number of hospital super-peer agents in an HBN to reduce the redundancy of big data storage. Furthermore, by using a temporary and permanent block scheme, EHRs can be efficiently published in HBNs. Finally, in previous work [7], new access control policies need to be established and approved after the search results are returned. In this work, we require that access control policies be established prior to the doctor's visits. Thus, the search and retrieval steps of EHRs can be combined to achieve an efficient information search and retrieval process.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 presents a cloud-based hierarchical consortium blockchain framework. Section 4 introduces the block structures and the processes of generating and publishing new blocks in different blockchains. Section 5 describes the search and retrieval process of EHRs in details. Section 6 presents the case studies and their analysis results. Section 7 concludes the paper and mentions future work.

2. Related Work

There are various studies and explorations on blockchain technology to develop a decentralized storage and sharing system for the healthcare sector [3]-[5]. Blockchain technology has been shown to be a viable technology as it allows sharing of medical data among approved healthcare providers while maintaining patient privacy [8]. Further research has also addressed the

challenge of storing big data such as images and videos in the blockchain; however, these studies have typically utilized off-chain approaches to store big data, rather than on-chain solutions. In [9], the authors proposed a storage model based on blockchain and InterPlanetary File System (IPFS) to store transactions efficiently in blockchain. In their design, the actual patient reports are stored in distributed off-chain storage using IPFS, while the blockchain stores only hash values of the reports, thereby reducing the overall block size in the blockchain. In [10], the authors developed a decentralized and permissible blockchain-based application for storing and accessing satellite task-scheduling schemes using the Hyperledger Fabric framework. They used IPFS for off-chain storage to reduce the asset size of the transaction and increase the transaction throughput of the network. In [11], the authors introduced a video surveillance storage and sharing system using blockchain technology. In their approach, videos received from the camera are encrypted and stored off-chain through distributed IPFS system with their metadata stored in the blockchain. Some researchers also proposed a secure data sharing solution for sensitive financial data using blockchain and proxy re-encryption technology [12]. Access control rules, hash values, and storage addresses of financial data are stored in the blockchain, while the actual financial data are stored off-chain in distributed databases. In [13], the authors proposed a blockchain framework using an attribute-based cryptosystem for the development of a secure EHR storage and sharing system. In their approach, large-scale medical data are stored in the cloud and the blockchain stores only the metadata of EHRs. Unlike these approaches, our cloud-based blockchain solution enables all healthcare data, including multimedia files, to be stored in the blockchains. Thus, our on-chain data storage approach provides the benefits of a complete blockchain storage solution in terms of data immutability, integrity and availability.

There are other studies focusing on the design of new blockchain architecture, which are summarized below. In [14], the authors proposed Fortified-Chain, a decentralized EHR and blockchain-based distributed data storage system (DDSS). They designed a global DDSS network that facilitates communications between local DDSS networks consisting of hospitals and third-party health services that store patient medical data. In [15], the authors developed a simplified version of a scalable blockchain architecture for sharing EHRs among patients, healthcare professionals and health institutions. In their approach, each health facility implements a local blockchain network connected to a global blockchain system to allow interaction between different health institutions. Some researchers also studied and introduced a blockchainless approach based on directed acyclic graphs (DAGs) for trusted public construction bidding to ensure fairness in the bidding process [16]. The DAG-based approach differs from the traditional blockchain because in the chainless approach, the DAG links the transaction containing its parents, documents, and a list of transaction signatures to other transactions through a less complex verification process. In a more recent effort, researchers designed a Compacted DAG-based blockchain protocol (CoDAG), used in the field of Industrial Internet of Things (IIoT) [17]. They developed protocols and algorithms to secure the network and confirm transactions within a specified time. The aforementioned blockchain-based approaches either use off-chain storage, e.g.,

[14] and [15], or do not address scalability issues, e.g., [16] and [17]. Unlike the above approaches, our novel cloud-based hierarchical blockchain architecture not only supports on-chain storage of big data, but also allows interaction and data sharing between peers located in different cities and states. Since EHRs from different peers cities are stored in different HBNs, our cloud-based hierarchical blockchain approach provides a scalable solution for storing sensitive information and big data in a nationwide network of connected consortium blockchains.

Previous research efforts on implementing access control mechanisms in blockchain networks have focused on preventing unauthorized access to confidential data stored in the blockchains. In [4], the authors proposed MeDShare, a blockchain-based system that enables peer-to-peer medical data sharing in a trustless environment. They used smart contracts and access control mechanisms to monitor and track the behavior of storing data in the blockchain. If any form of data permission violation is detected, the system revokes the access rights of the offending user. In [18], the authors proposed a Blockchain-as-a-Service based solution for Health Information Exchange (BaaS-HIE) activities to deal with security issues including patient privacy, integrity of medical records, and fine-grained access control. Their approach involves the use of a private blockchain based on the Ethereum protocol and smart contracts as access control management for medical records. In a similar way, other researchers designed an access control mechanism on managing user access to ensure efficient and secure sharing of EHRs on mobile devices by leveraging smart contracts on the Ethereum blockchain [19]. In [20], the authors proposed the use of blockchain and edge nodes to facilitate attribute-based access control and storage of EHR data. They used smart contracts to enforce access control of EHR data stored in off-chain edge nodes. In their subsequent work, encryption for data stored at the edge nodes was further developed [21]. The multi-authority attribute-based encryption (ABE) scheme and attribute-based multi-signature (ABMS) scheme were used to encrypt the EHR data stored at the edge nodes and verify users' signatures, respectively. In contrast to the above work, our approach involves the implementation of different scopes of role-based access control (RBAC) policies that restrict user access to various healthcare facilities in different cities and states. We define three layers of the networks, each implementing its own RBAC policies – local hospital-wide policies, city-wide policies, and statewide policies, respectively. As a result, our approach provides a more comprehensive and reliable mechanism than other methods because it is designed to work in a much larger environment.

3. A Framework for Hierarchical Blockchains

The framework for cloud-based hierarchical consortium blockchain networks consists of three layers. As shown in Figure 1, these layers are the *Hospital Layer*, the *City Layer*, and the *State Layer*, which contain multiple HBNs, multiple CBNs, and an SBN, respectively. An HBN in the hospital layer covers multiple hospitals from the same city or local area, represented by hospital super-peer agents β_{HOSs} or hospital regular-peer agents β_{HREPs} . A CBN in the city layer covers multiple cities from the same state. A city super-peer agent β_{CIT} acts as a representative of a city and a network regulator for the HBN belonging to the city. Finally, an SBN in the state layer covers all states of a country. A state super-

peer agent β_{STA} acts as a representative of a state and a network regulator for the CBN belonging to the state.

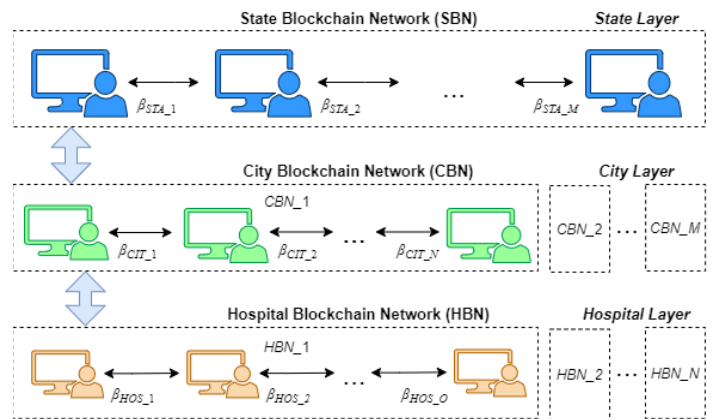


Figure 1: The Architecture of Cloud-Based Hierarchical Consortium Blockchains

To enable big data storage in blockchain networks, a cloud-based blockchain scheme is implemented in the hospital layer, i.e., HBNs. Unlike previous work [5], which requires all participating hospitals in an HBN to implement cloud-based blockchains, in this study, the HBN consists of three types of agents, namely the hospital super-peer agents β_{HOSs} , representing designated hospitals, hospital regular-peer agents β_{HREPs} , representing general hospitals, and regular-peer agents β_{REPs} , representing end users including doctors, nurses, and patients. We define general hospitals as those that do not have the required infrastructure to implement cloud-based blockchain storage or choose not to do so. Figure 2 shows an example of an HBN where hospital A and B are designated hospitals that offer private cloud services, while hospital C is a general hospital that do not provide such services.

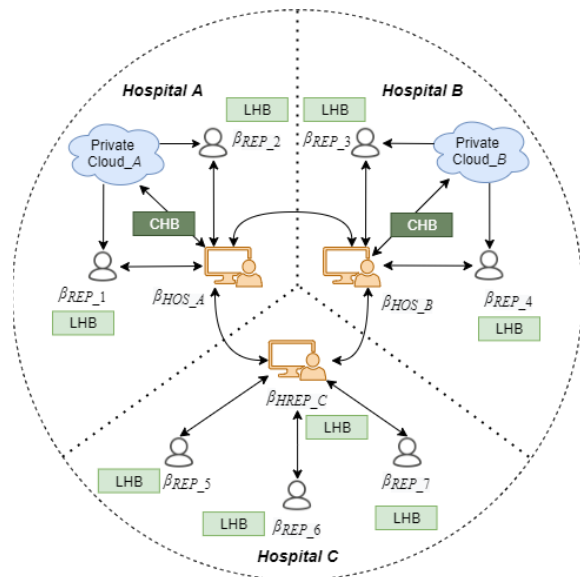


Figure 2: The Relationships Between Participants in an HBN

As shown in Figure 2, a cloud-based hospital blockchain (CHB) is implemented in the designated hospitals with their private clouds. A CHB is managed by a hospital super-peer agent β_{HOS} and stores all data, including EHRs in multimedia file format. To avoid excessive redundancy of big data in an HBN, we limit the number of hospital super peer agents in an HBN to no more

than 6-10. A lite hospital blockchain (LHB) is implemented on the server of a general hospital, managed by a hospital regular-peer agent β_{HREP} , or on the local machine of an end user, managed by a regular-peer agent β_{REP} . An LHB stores all data in its corresponding CHB, except for big data such as multimedia files, for which only their metadata are stored in the LHB. Access to confidential data, i.e., a patient's EHRs, stored in a CHB is managed by a hospital super-peer agent β_{HOS} , while access to confidential data stored in an LHB is managed by either a hospital super-peer agent β_{HOS} or a hospital regular-peer agent β_{HREP} .

Figure 3 shows the general blockchain structure and the similarity between CHB and LHB. Let the length of a LHB and its corresponding CHB be h . A cloud-based block CB_i and a lite block LB_i , where $1 \leq i \leq h$, contain the same information except for the multimedia files. This scheme allows an end user or a general hospital to use the metadata stored in its LHB to submit a request to a relevant hospital super-peer agent β_{HOS} through its regular-peer agent β_{REP} or hospital regular-peer agent β_{HREP} and retrieve the corresponding multimedia files stored in the CHB.

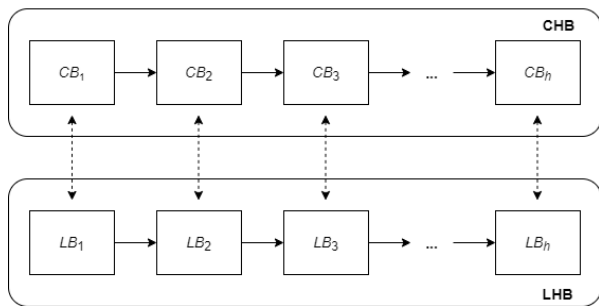


Figure 3: The General Blockchain Structure of a CHB and an LHB

To support the timely publication of EHRs in an HBN, we introduce a temporary and permanent block scheme based on earlier work [6]. Due to the need to publish EHRs, including their associated multimedia files in a timely manner, temporary blocks are only included in the hospital layer of our cloud-based hierarchical blockchain networks; however, they are not required in the city and state layers, as access control policies and access records do not need to be published immediately. Figure 4 shows an example of a CHB with temporary and permanent blocks.

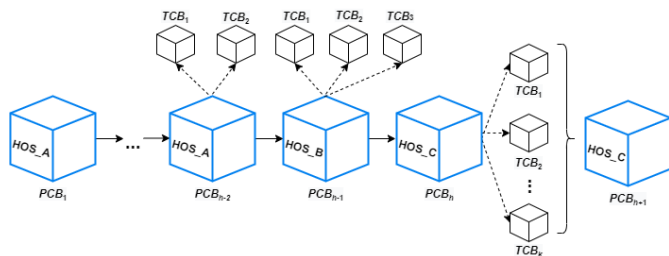


Figure 4: A CHB with Temporary and Permanent Blocks

As shown in Figure 4, a block PCB_i , where i is the height of the block in the blockchain, is a cloud-based permanent block. To support efficient information retrieval, a PCB contains only EHRs and other related records from the same hospital. On the other hand, a block TCB_j , where j denotes the order in which the temporary blocks are attached to a PCB according to their publishing time, is a cloud-based temporary block. To support the timely publication of block records, a TCB stores only one record

from a hospital and must be attached to the latest PCB published for the same hospital. As shown in the figure, the latest PCB for a hospital can be attached by multiple $TCBs$ that are numbered in the order in which they appear. Note that the LHB corresponding to a CHB shares the same structure but have different notations, i.e., permanent lite block PLB and temporary lite block TLB .

Since a TCB contains only one block record, whenever a block record is generated and submitted to a super-peer agent, the agent can immediately publish the block record to the blockchain as a temporary block through the temporary block generation process. Meanwhile, a permanent block stores multiple block records as in a typical blockchain. Once enough $TCBs$ are generated and published to the blockchain by a hospital super-peer agent, the agent can consolidate them into a new PCB through the permanent block generation process. For example, as in Figure 4, when the number or total size of $TCBs$ attached to PCB_h reaches a threshold, the agent β_{HOS_C} merges the list of $TCBs$ and forms a new cloud-based permanent block PCB_{h+1} for publishing. When PCB_{h+1} is published, all $TCBs$ attached to PCB_h are removed from the blockchain. Note that other $TCBs$ that are attached to $PCBs$ other than PCB_h will remain in the blockchain until they are merged into new $PCBs$.

4. Publication of New Blocks in the Blockchain Networks

An HBN, a CBN or the SBN maintains its own blockchain, namely hospital blockchain, city blockchain or state blockchain, respectively. Blockchain networks of the same type, such as two HBNs, are independent, but they can communicate through a higher-level blockchain network, e.g., a CBN if the two HBNs belong to the same state, or the SBN if the two HBNs are in different states. Hospital, city, and state blockchains can store different types of block records for different purposes. In this section, we describe the types of block records used in different blockchains and the procedures for generating and publishing new blocks in different types of blockchains.

4.1. Hospital Block and its Block Record Types

There are four different types of block records that can be used in a CHB or an LHB, namely HR_{UPR} , HR_{ACP} , HR_{MER} , and HR_{AR} . To simplify matters, we define a hospital blockchain as a general term that can refer to a CHB or an LHB. We now describe the four types of block records as follows.

- HR_{UPR} is a record that stores the account information and user profile of an end user, represented by regular-peer agent β_{REP} . An HR_{UPR} is defined as a 6-tuple (I, N, R, U, S, T) , where I is the identification of β_{REP} in the HBN; N is the full name of β_{REP} ; R , U and S are β_{REP} 's private key, public key and secret symmetric key, respectively; and T is the timestamp when the HR_{UPR} is created. Whenever a new user joins the HBN or an existing user's profile is updated, a new HR_{UPR} is created.
- HR_{ACP} is a record that stores access control policies and is used to conduct permission checks on requests to access EHRs stored at hospitals within the same city. An HR_{ACP} is defined as a triple (P, H, T) , where P is a set of policies; H is a set of hospital where the policies are executed; and T is the timestamp when the policies are created.
- HR_{AR} is a record that stores information on access requests to a patient's EHRs stored at hospitals within the same city where

the patient resides. HR_{AR} is created as a log of access requests for accountability purposes. An HR_{AR} is defined as 4-tuple (N, D, O, T) , where N is the request number; D is the detail of the request; O is the outcome of the request; and T is the timestamp when the request is created.

- HR_{MER} is a record that stores medical information, including patient reports and metadata for any related multimedia files generated after a doctor’s visit. An HR_{MER} is defined as 5-tuple (I, H, X, M, T) , where I are the identifications of all peers involved in the doctor’s visit, including the patient, the nurse and the doctor; H is the name of the hospital where the patient visited; X includes a summary of the visit and any text-based medical data; M is the metadata of any multimedia files generated after the doctor’s visit; and T is the timestamp when the HR_{MER} record is created.

Since both permanent and temporary blocks in a CHB may contain multimedia files, the blocks PCB and TCB consist of two major components: the block component and the multimedia file component. Figure 5 shows the block structure of a new temporary cloud-based block TCB_j with three sections in the block component and one section in the multimedia file component. These are header, hospital block records, verification information, and multimedia files in an EHR.

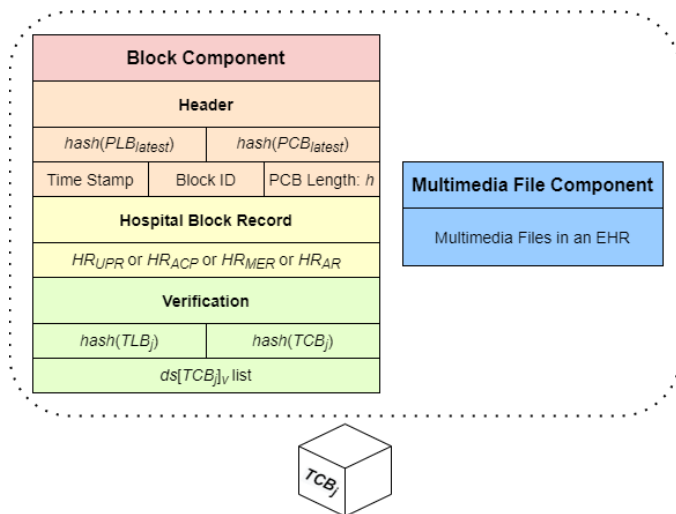


Figure 5: The Structure of a New Temporary Cloud-Based Block TCB_j

As shown in Figure 5, the header section contains the hash values of the latest PLB and PCB , published by the same hospital super-peer agent who generates TCB_j , the timestamp when TCB_j was created, the block ID, and the length h of the current blockchain. The hospital block records section contains only a single block record of HR_{UPR} , HR_{ACP} , HR_{MER} , or HR_{AR} as TCB_j is a temporary block. Consequently, the multimedia file section can only store multimedia files from one doctor’s visit, if any, while their metadata is recorded and stored in the relevant HR_{MER} in the hospital block record section. Lastly, the verification section contains the hash values of the current block, including the hash value of the header and hospital block records, denoted as $hash(TLB_j)$, and the hash value of the header, hospital block records and the multimedia files, denoted as $hash(TCB_j)$. The verification section also contains a list of digital signatures $ds[TCB_j]_v$, where each peer v is an agent β_{HOS} who approves TCB_j during the temporary block consensus process. Note that the

structure of the temporary lite block TLB_j is similar to that of TCB_j but does not include the multimedia file component.

Figure 6 shows the block structure of a new permanent cloud-based block PCB_{h+1} . The block structure of PCB is similar to that of TCB , but a PCB can accommodate multiple block records and EHRs in its hospital block records section and multimedia file component, respectively.

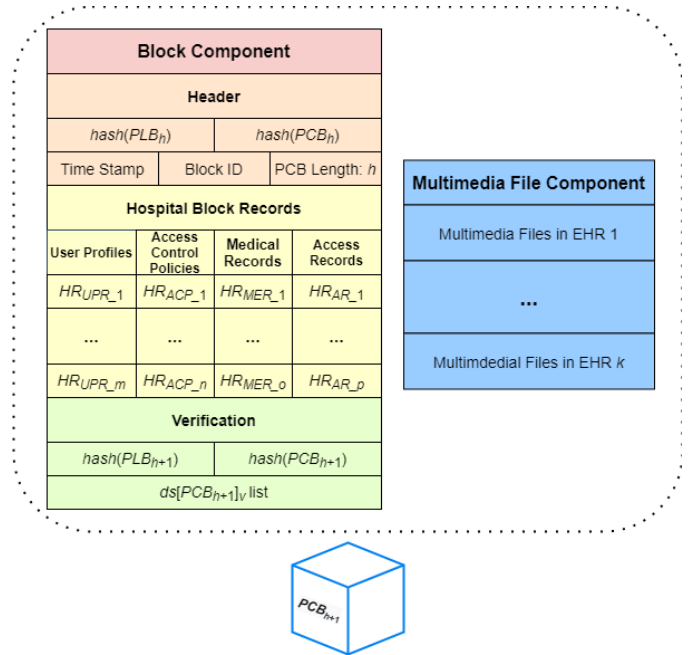


Figure 6: The Structure of a New Permanent Cloud-Based Block PCB_{h+1}

In a PCB , the verification section contains the hash values of the permanent lite block PLB_{h+1} and the permanent cloud-based block PCB_{h+1} . It also contains a list of digital signatures $ds[PCB_{h+1}]_v$, where each peer v is an agent β_{HOS} , who approves PCB_{h+1} during the permanent block consensus process. Note that while a new TCB is attached to the latest PCB , published by the hospital super-peer agent who generates the TCB , a new PCB must be attached to the last PCB of the cloud-based blockchain, i.e., PCB_h , where h is the height of the current blockchain. As with temporary blocks, the structure of a permanent lite block PLB_{h+1} is similar to that of PCB_{h+1} except for the inclusion of the multimedia file component in PCB_{h+1} .

4.2. Hospital Temporary and Permanent Block Generation

Let hospital super-peer agent $\beta_{HOS-\psi}$ be the one who creates a new temporary cloud-based block TCB_j . Algorithm 1 shows how the new block TCB_j is generated by agent $\beta_{HOS-\psi}$. According to the algorithm, agent $\beta_{HOS-\psi}$ first creates an empty temporary cloud-based block TCB_j . All attributes in TCB_j ’s header section are then created and added. These include the hash values of the latest permanent blocks, previously published by $\beta_{HOS-\psi}$, i.e., $hash(PCB_{latest})$ and $hash(PLB_{latest})$, the timestamp when TCB_j is created, TCB_j ’s block ID, and the blockchain length h . After that, $\beta_{HOS-\psi}$ processes the hospital block record φ given in the input list accordingly. If φ is an HR_{UPR} and $\varphi.S$ is null, it indicates that φ records the account information and user profile of a new end user. In this case, a secret symmetric key is automatically generated and added it to $\varphi.S$. Then φ is encrypted using the public key of β_{HOS} .

ψ and added to the hospital block record section of TCB_j . If ϕ is a medical block record HR_{MER} and a list ρ of multimedia files is included, $\beta_{HOS-\psi}$ encrypts the files in ρ using the associated patient's secret symmetric key retrieved from the patient's latest HR_{UPR} . The encrypted files are then added to the multimedia file component of TCB_j . The metadata of the encrypted files are also recorded and added to $\phi.M$ of the medical block record. Finally, ϕ itself is encrypted, except for $\phi.M$, before it is added to the hospital block record section of TCB_j . Note that if ϕ is an HR_{ACP} or HR_{AR} , it is added directly to the hospital block record section of TCB_j in plaintext. Once the header section and hospital block record section are established, $\beta_{HOS-\psi}$ calculates the hash values $hash(TLB_j)$ and $hash(TCB_j)$, as well as the digital signature $ds[TCB_j]_\psi$ using $hash(TCB_j)$. All these elements are then added to the verification section of TCB_j . Note that while not shown in Algorithm 1, a new temporary lite block TLB_j can be created by simply removing the multimedia file component from TCB_j .

Algorithm 1: Generating a New Temporary Block TCB_j

Input: A hospital block record ϕ containing record HR_{UPR} , HR_{ACP} , HR_{AR} , or HR_{MER} , and an optional list ρ of multimedia files.

Output: A new temporary cloud-based block TCB_j .

1. Create an empty temporary cloud-based block TCB_j
 2. Verify and add $hash(PCB_{latest})$, $hash(PLB_{latest})$, time stamp, block ID and current blockchain length h to the header section of TCB_j
 3. **if** ϕ is an HR_{UPR} and $\phi.S$ is *null* // indicates a new end user
 4. Generate a secret symmetric key, add it to $\phi.S$, and encrypt ϕ
 5. **else if** ϕ is an HR_{MER} and ρ is not empty
 6. Encrypt the multimedia files in ρ
 7. Add the encrypted files to the multimedia file section of TCB_j
 8. Add the metadata of ρ to $\phi.M$ and encrypt ϕ , except for $\phi.M$
 9. Add ϕ to the hospital block record section of TCB_j
 10. Calculate the hash values $hash(TCB_j)$ and $hash(TLB_j)$
 11. Add the hash values to the verification section of TCB_j
 12. Create digital signature $ds[TCB_j]_\psi$ using $hash(TCB_j)$
 13. Add $ds[TCB_j]_\psi$ to the $ds[TCB_j]_\psi$ list in the verification section
 14. **return** TCB_j
-

Once enough $TCBs$ are generated and published to the blockchain by $\beta_{HOS-\psi}$, the $TCBs$ can be consolidated into a new permanent block PCB through a permanent block generation process. Algorithm 2 shows how a new permanent cloud-based block PCB_{h+1} is generated by agent $\beta_{HOS-\psi}$. According to the algorithm, agent $\beta_{HOS-\psi}$ first creates an empty permanent cloud-based block PCB_{h+1} . All attributes in PCB_{h+1} 's header section, including $hash(PCB_h)$ and $hash(PLB_h)$, the timestamp, the block ID, and the blockchain length h , are then created and added. For each temporary block τ in the temporary block list Ξ , $\beta_{HOS-\psi}$ verifies it using information stored in τ 's header and the verification section and transfers all relevant information from the hospital block record section of τ to the hospital block records section of PCB_{h+1} as a new block record. If τ contains a block record HR_{MER} and a list of encrypted multimedia files ρ , $\beta_{HOS-\psi}$ moves files in ρ to the multimedia file component of PCB_{h+1} and adds the metadata of ρ to the relevant block record $HR_{MER}.M$. This ensures that all previously stored information in the temporary blocks from the list Ξ is transferred to PCB_{h+1} . Finally, $\beta_{HOS-\psi}$ calculates the hash values $hash(PCB_{h+1})$ and $hash(PLB_{h+1})$, as well as the digital signature $ds[PCB_{h+1}]_\psi$ using $hash(PCB_{h+1})$. All these elements are then added to the verification section of PCB_{h+1} .

Similar to the generation of TLB_j , a new permanent lite block PLB_{h+1} can be created by simply removing the multimedia file component from PCB_{h+1} .

Algorithm 2: Generating a New Permanent Block PCB_{h+1}

Input: A list of blocks Ξ containing k temporary cloud-based blocks.

Output: A new permanent cloud-based block PCB_{h+1} .

1. Create an empty permanent cloud-based block PCB_{h+1}
 2. Verify and add $hash(PCB_h)$, $hash(PLB_h)$, time stamp, block ID, and current blockchain length h to the header section of PCB_{h+1}
 3. **for** each temporary cloud-based block τ in Ξ
 4. Verify τ and add all relevant parts from the hospital block record section in τ to the hospital block records section in PCB_{h+1}
 5. **if** τ contains HR_{MER} and a list of encrypted multimedia files ρ
 6. Add files in ρ to PCB_{h+1} 's multimedia file component
 7. Add the metadata of ρ to the corresponding $HR_{MER}.M$
 8. Calculate hash values $hash(PCB_{h+1})$ and $hash(PLB_{h+1})$
 9. Add the hash values to the verification section of PCB_{h+1}
 10. Create digital signature $ds[PCB_{h+1}]_\psi$ using $hash(PCB_{h+1})$
 11. Add $ds[PCB_{h+1}]_\psi$ to the $ds[PCB_{h+1}]_\psi$ list in the verification section
 12. **return** PCB_{h+1}
-

4.3. City and State Block and their Block Record Types

Unlike the hospital blockchain, there is only one variant of the city and state blockchains due to the absence of regular peers and big data. Thus, implementing cloud-based versions of city and state blockchains is not necessary. For city blockchain, there are two types of block records that can be stored in the blockchain. These are city-wide record for access control policies CR_{ACP} and city-wide access record CR_{AR} . CR_{ACP} is a record that stores the access control policies implemented in a CBN and enforced by the relevant city super-peer agent β_{CIT} . CR_{ACP} has the same structure as HR_{ACP} , except that the $CR_{ACP}.H$ contains additional information such as the names of cities and hospitals where the policies are enforced. CR_{ACP} is created to check any requests regarding access to patient EHRs stored in HBNs across cities within the same state. On the other hand, CR_{AR} is a record that stores information on access requests to patient EHRs in hospitals across cities within the same state. The structure of a CR_{AR} is also similar to that of an HR_{AR} .

For state blockchain, a state block shares the same structure as that of a city block and stores statewide records for access control policies SR_{ACP} and statewide access record SR_{AR} . SR_{ACP} is a record that stores the access control policies implemented in the SBN and enforced by the relevant state super-peer agent β_{STA} . SR_{ACP} is established to check for any access requests to patient EHRs stored in HBNs across states; while SR_{AR} is a record that stores information on access requests to patient EHRs in hospitals across states. Figure 7 shows the structure of a new city or state block B_{h+1} in a city or state blockchain. From the figure, we can see that block B_{h+1} consists of only one component as city and state blockchains do not store EHRs. There are three sections present in block B_{h+1} , namely header, state or city block records, and the verification section. The header section contains the previous city or state block's hash value $hash(B_h)$, the timestamp when B_{h+1} is created, the block ID of B_{h+1} , and the current blockchain length h . The city or state records section contains a list of block record CR_{ACP} and/or CR_{AR} , or SR_{ACP} and/or SR_{AR} , respectively. The verification section contains the hash values of B_{h+1} and a list of

digital signatures $ds[B_{h+1}]_v$, where each peer v is a city or state super-peer agent who approves B_{h+1} during the consensus process.

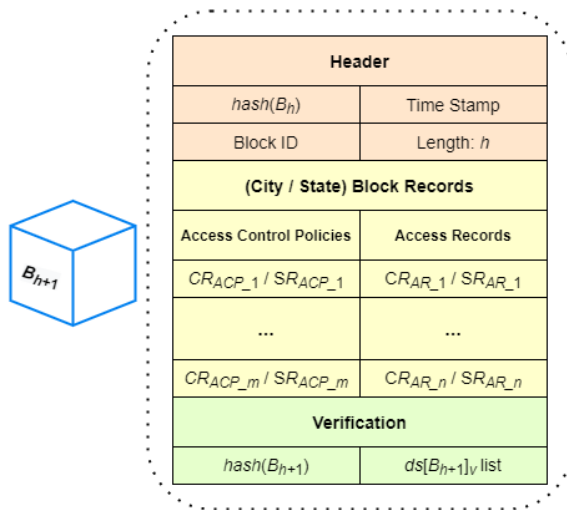


Figure 7: The Structure of a New City or State Block B_{h+1}

4.4. City and State Block Generation

The process of generating a new city or state block is similar to that of generating a lite hospital block, although it is simpler due to the absence of big data and end users. Let the city super-peer agent $\beta_{CIT-\psi}$ be the one who creates the new city block B_{h+1} . Algorithm 3 shows the procedure for generating B_{h+1} , which is then approved and added to the city blockchain through a city block consensus process.

Algorithm 3: Generating a New City Block B_{h+1}

Input: A list of city block records Φ containing CR_{ACP} and/or CR_{AR}
Output: A new city block B_{h+1}

1. Create an empty city block B_{h+1}
2. Verify and add $hash(B_h)$, time stamp, block ID, and current blockchain length h to the header section of B_{h+1}
3. **for** each record φ in the list Φ of city block records
4. **if** φ is an CR_{ACP}
5. Add φ to the city block records section of B_{h+1}
6. **else**
7. Encrypt φ and add it to the city block records section
8. Calculate $hash(B_{h+1})$ and add it to the verification section
9. Create digital signature $ds[B_h]_\psi$ using $hash(B_{h+1})$
10. Add $ds[B_h]_\psi$ to the $ds[B_h]_v$ list in the verification section
11. **return** B_{h+1}

According to the algorithm, agent $\beta_{CIT-\psi}$ first creates an empty city block B_{h+1} . All attributes in B_{h+1} 's header section are then created and added. These include the previous block's hash value $hash(B_h)$, the timestamp when B_{h+1} is created, the block ID, and the blockchain length h . After that, $\beta_{CIT-\psi}$ processes all records in the city block record list Φ accordingly before they are added to the city block records section of B_{h+1} . If a city block record φ is a CR_{ACP} , it is simply added to B_{h+1} 's city block records section without being encrypted. Afterwards, $\beta_{CIT-\psi}$ calculates $hash(B_{h+1})$ and $ds[B_{h+1}]_\psi$ before adding them to the verification section of B_{h+1} . Note that the algorithm for generating a new state block is similar to Algorithm 3 due to the shared structure of the city and state blocks.

4.5. Temporary and Permanent Block Consensus Process

In our approach, we implemented a simple majority vote consensus mechanism for publishing new hospital, city, and state blocks. The consensus processes implemented in HBN, CBN and SBN function similarly. Let λ be the total number of super-peer agents from a blockchain network who participate in a consensus process. The block announcer, the super-peer agent who is responsible for initiating the consensus process, must broadcast the new block to other super-peer agents in the network and gather at least $\lambda/2$ approvals from super-peer agents within the same blockchain network.

Figure 8 shows a general illustration of the consensus process for approving a new temporary block in an HBN. From the figure, we can see that the temporary block consensus process consists of 7 steps. The first step is the announcement of a newly created temporary block TCB_j by the block announcer β_{HOS_A} to the super-peer agents of other hospitals in the network. To simplify matters, we show only one such agent in the figure, i.e., β_{HOS_B} . Note that hospital regular-peer agent β_{HREP_C} does not participate in the consensus process as it does not have direct access to the CHB. Once the announcement is broadcast and received, β_{HOS_B} retrieves TCB_j from the block announcer in step 2. After that, in step 3, β_{HOS_B} verifies the validity of TCB_j by checking the integrity of TCB_j and the digital signature of the block announcer in the block. If TCB_j is considered valid, β_{HOS_B} creates its digital signature and sends it back to the block announcer as an approval vote in step 4. The block announcer waits for a certain amount of time in step 5 until either a timeout is reached, or a majority of approval votes are collected. All valid digital signatures are added to the digital signature list of $ds[TCB_j]_v$. If a majority vote is received by the block announcer β_{HOS_A} , block TCB_j is considered complete and can be added to the CHB. In this case, agent β_{HOS_A} notifies β_{HOS_B} that block TCB_j has successfully passed the consensus process in step 6. Finally, in step 7, each hospital super-peer agent with a completed TCB_j can generate a lite temporary block TLB_j and broadcast it to its respective regular-peer and hospital regular-peer agents for inclusion of TLB_j in their LHBs.

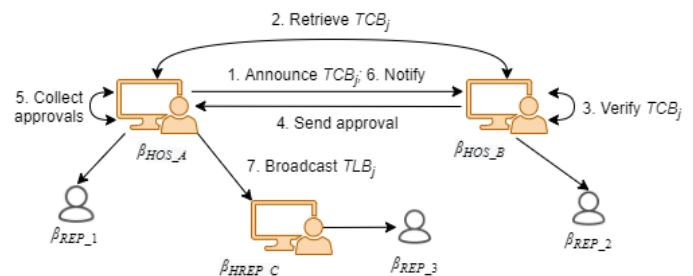


Figure 8: The Consensus Process for Approving a New Temporary Block

As more and more temporary blocks are added to the blockchain by a hospital super-peer agent β_{HOS} through the temporary block consensus process, agent β_{HOS} can decide to merge all its own published temporary blocks into one permanent block by initiating the permanent block consensus process. Note that when a permanent block consensus process is initiated, no other permanent block or temporary block consensus processes are allowed to occur at the same time and vice versa. The consensus process for approving a new permanent block in an HBN is similar to that for approving a new temporary block in an

HBN, depicted in Figure 8, but it requires the deletion of all temporary blocks that have been merged into a new permanent block in its last step. Finally, the consensus process for approving a new city or state block in a CBN or the SBN, respectively is also similar. More details can be found in a recent work [7].

5. The Search and Retrieval Processes for EHRs

There are numerous peers involved in the blockchain networks, playing different roles such as doctors, nurses and patients. Before retrieving EHRs from the blockchains, it is critical to assign appropriate permissions to each role to access the EHRs stored in the blockchains and protect them from unauthorized access [22]. In this section, we first describe the RBAC policies used in our approach, and then present our integrated search and retrieval process for EHRs.

5.1. Role-Based Access Control Policy

In our previous work, we implemented RBAC policies in an HBN as mandatory rules that specify which data in a blockchain can be accessed by participants based on their credentials [5]. With the introduction of hierarchical blockchain framework, RBAC policies are required to function effectively across all three layers of the blockchain networks. In other words, RBAC policies must be defined to grant access to a patient's EHRs across hospitals, cities, or states. These RBAC policies are stored in HR_{ACP} , CR_{ACP} and SR_{ACP} of a hospital blockchain, a city blockchain and a state blockchain, respectively. A patient is required to decide whether to allow or deny the sharing of their medical data with other hospitals within the city, state, or country prior to a doctor's visit. Any relevant access control policies are then created and added to the appropriate hospital, city, or state blockchains. A regular peer agent β_{REP} that represents an end user (e.g., a doctor), must seek permission from a hospital, city, or state super-peer agent for access to a patient's EHRs stored within hospitals either locally or across the country. An example policy H1 is shown below, which is stored as an HR_{ACP} in a hospital blockchain and enforced by the hospital super-peer agents within the corresponding HBN.

```

policy H1 {
  summary: Doctor D#111 from Hospital_1 is allowed access to
  Patient P#112's EHRs in Hospital_2.
  hospitals: Hospital_1; Hospital_2
  role: doctor (Doctor D#111), patient (Patient P#112)
  condition: doctor ∈ Hospital_1 && patient ∈ Hospital_2
  owners:  $\beta_{HOS-HOSPITAL_1}$ ;  $\beta_{HOS-HOSPITAL_2}$ 
  expiration: 01/01/2031
}
    
```

Access control policy H1 specifies that doctor D#111 is allowed to access patient P#112's EHRs in Hospital_2. Note that since a hospital access control policy HR_{ACP} specifies access rights within an HBN, we can safely assume that Hospital_1 and Hospital_2 are located in the same city. When doctor D#111 makes a request to access patient P#112's EHRs, both hospital super-peer agents $\beta_{HOS-HOSPITAL_1}$ and $\beta_{HOS-HOSPITAL_2}$ attempt to verify the request by checking policy H1 stored in their blockchains. If approved, doctor D#111 is granted access to patient P#112's EHRs maintained by Hospital_2. A city access control policy stored as a CR_{ACP} is similar to a hospital access control policy stored as an HR_{ACP} , but it must specify the cities where the hospitals are located because the hospitals belong to different cities within the same state;

otherwise, if the hospitals belong to the same city, the access control policy shall be recorded as an HR_{ACP} . An example policy C1 is shown below, which can be stored in a city blockchain as a CR_{ACP} and enforced by city super-peer agents in CBNs for the HBNs under their jurisdiction.

```

policy C1 {
  summary: Doctor D#111 from Hospital_1 (City_1) is allowed access to
  Patient P#112's EHRs in Hospital_3 (City_3).
  hospitals: City_1.Hospital_1; City_3.Hospital_3
  role: doctor (Doctor D#111), patient (Patient P#112)
  condition: doctor ∈ City_1.Hospital_1 && patient ∈ City_3.Hospital_3
  owners:  $\beta_{CIT-City_1}$ ;  $\beta_{CIT-City_3}$ 
  expiration: 02/02/2032
}
    
```

Access control policy C1 specifies that doctor D#111 from Hospital_1 in City_1 is allowed to access patient P#112's EHRs at Hospital_3 in City_3. Different from policy H1, when doctor D#111 makes a request to access patient P#112's EHRs located in a different city, both hospital super-peer agents $\beta_{HOS-HOSPITAL_1}$ and $\beta_{HOS-HOSPITAL_3}$ forward the request to their city super-peer agents $\beta_{CIT-City_1}$ and $\beta_{CIT-City_3}$ to check against policy C1 stored in their city blockchains. If approved, doctor D#111 is granted access to patient P#112's EHRs at Hospital_3 in City_3.

A state access control policy stored as an SR_{ACP} is similar to a city access control policy stored as a CR_{ACP} , but it must specify both the cities and the states where the hospitals are located. An example policy S1 is shown below, which can be stored as an SR_{ACP} in a state blockchain and enforced by state super-peer agents in the SBN for CBNs and HBNs under their jurisdiction.

```

policy S1 {
  summary: Doctor D#111 from Hospital_1 (City_1, State_1) is allowed access
  to Patient P #112's EHRs in Hospital_4 (City_4, State_4).
  hospitals: State_1.City_1.Hospital_1; State_4.City_4.Hospital_4
  role: doctor (Doctor D#111), patient (Patient P#112)
  condition: doctor ∈ State_1.City_1.Hospital_1 &&
  patient ∈ State_4.City_4.Hospital_4
  owners:  $\beta_{STA-State_1}$ ;  $\beta_{STA-State_4}$ 
  expiration: 03/03/2033
}
    
```

Access control policy S1 specifies that doctor D#111 from Hospital_1 (City_1, State_1) is allowed to access patient P#112's EHRs stored in Hospital_4 (City_4, State_4). In a similar nature to policy C1, when doctor D#111 makes a request to access patient P#112's EHRs located in a different state, both hospital super-peer agents $\beta_{HOS-HOSPITAL_1}$ and $\beta_{HOS-HOSPITAL_4}$ forward the request to their state super-peer agents $\beta_{STA-State_1}$ and $\beta_{STA-State_4}$, through their city super-peer agents, $\beta_{CIT-City_1}$ and $\beta_{CIT-City_4}$. The request is then checked against policy S1 stored in their state blockchains. If approved, doctor D#111 is granted access to patient P#112's EHRs from Hospital_4 (City_4, State_4).

Note that to support efficient access authorization and avoid duplication of an access control policy across multiple access control policy records, access control policies are no longer encrypted as in our previous work [7]. For more examples of access control policies at hospital, city and state levels, refer to earlier work [5], [7].

5.2. Integrated Search and Retrieval of EHRs

Once the required access control policies have been created and stored in the relevant hospital, city and state blockchains, the

associated data can now be opened and shared with other hospitals across the country. This data sharing is supported by an integrated EHRs search and retrieval process that enables those with the proper authorization to retrieve all EHRs of a patient from any hospitals, regardless of which HBNs they participate in. This process involves all three layers of our hierarchical blockchain framework, as search requests are forwarded and executed concurrently across all super-peer agents in the hierarchical network structure. The concurrent search and retrieval process is defined by three procedures, which are searching and retrieving EHRs across hospitals within the same city, searching and retrieving EHRs across cities within the same state, and searching and retrieving EHRs across states within a country. We now describe each of the three procedures as follows.

The procedure of searching and retrieving EHRs across hospitals within the same city is presented in Algorithm 4. The algorithm is initiated by a hospital super-peer agent β_{HOS} on behalf of its end user (e.g., a doctor) to search and retrieve patient p 's EHRs from other hospitals within the same city (i.e., within the same HBN). Agent β_{HOS} sends this request to its city super-peer agent β_{CIT} to start the process.

Algorithm 4: Searching and Retrieving a Patient's EHRs from All Hospitals within the Same City by a City Super-Peer Agent β_{CIT}

Input: A retrieval request for hospitals containing patient p 's EHRs
Output: A list of links to patient p 's EHRs

1. Let ρ_h_list be the list of hospital super-peers under β_{CIT} 's jurisdiction
 2. Let η_{ehr_hlist} be an empty list of links to EHRs; $nResponse = 0$
 3. **for** each γ_h in ρ_h_list
 4. forward the retrieval request to γ_h asynchronously, which invokes a search process at hospital h based on the established policies
 5. **while** (not timeout) or $nResponse \neq |\rho_h_list|$
 6. **if** γ_h returns a link to p 's EHRs
 7. add the link to the list η_{ehr_hlist} ; $nResponse++$
 8. **else** $nResponse++$; **continue** // γ_h returns no link to EHRs
 9. **return** the list η_{ehr_hlist}
-

According to the algorithm, agent β_{CIT} sends concurrent requests in its HBN to all hospital super-peer agents under its jurisdiction and waits for responses or until the timeout. Each hospital super-peer agent γ_h who receives this request will perform a permission checking based on the established access control policies stored as HR_{ACP} in its hospital blockchain. If valid, γ_h creates a link that allows access to patient p 's EHRs and sends it back to β_{CIT} . If β_{CIT} receives a response from γ_h with this link, the link is added to list η_{ehr_hlist} ; otherwise, β_{CIT} continues to wait. When all hospital super-peer agents have responded or timed out, the list η_{ehr_hlist} is returned and sent back to β_{HOS} . Upon receiving η_{ehr_hlist} , β_{HOS} can then use the links to access and retrieve patient p 's EHRs on behalf of the end user.

The procedure of searching and retrieving EHRs across cities within the same state, is presented in Algorithm 5. Similar to the procedure of searching and retrieving EHRs across hospitals within the same city, Algorithm 5 is initiated by a hospital super-peer agent β_{HOS} on behalf of its end user (e.g., a doctor) to search and retrieve patient p 's EHRs from hospitals in different cities within the same state (i.e., within different HBNs connected under the same CBN). Agent β_{HOS} sends this request to its city super-peer agent β_{CIT} , who forwards it to its state super-peer agent β_{STA}

to start the process. According to the algorithm, agent β_{STA} sends concurrent requests to all city super-peer agents under its jurisdiction in its CBN and waits for responses from them or until it times out. Upon receiving the search request, each city super-peer agent γ_c performs a permission check based on the access control policies stored as CR_{ACP} in its city blockchain. If valid, γ_c executes Algorithm 4 to forward the search and retrieval requests to the hospital super-peer agents under its jurisdiction. If γ_c returns a list η_{ehr_hlist} containing links to p 's EHRs, η_{ehr_hlist} is appended to the list η_{ehr_clist} ; otherwise, β_{STA} continues to wait. When all city super-peer agents have responded or it times out, the list η_{ehr_clist} is returned and sent back to β_{CIT} , who further sends it back to β_{HOS} . Upon receiving η_{ehr_clist} , β_{HOS} can then use the links to access and retrieve patient p 's EHRs on behalf of the end user.

Algorithm 5: Searching and Retrieving a Patient's EHRs from All Hospitals within the Same State by a State Super-Peer Agent β_{STA}

Input: A retrieval request for hospitals containing patient p 's EHRs
Output: A list of links that can be used to access patient p 's EHRs

1. Let ρ_c_list be the list of city super peers under β_{STA} 's jurisdiction
 2. Let η_{ehr_clist} be an empty list of links to EHRs; $nResponse = 0$
 3. **for** each γ_c in ρ_c_list
 4. forward the retrieval request to γ_c asynchronously, which invokes Algorithm 4 to search in city c based on the established policies
 5. **while** (not timeout) or $nResponse \neq |\rho_c_list|$
 6. **if** γ_c returns η_{ehr_hlist} that contains links to p 's EHRs
 7. append η_{ehr_hlist} to η_{ehr_clist} ; $nResponse++$
 8. **else** $nResponse++$; **continue** // γ_c returns an empty list
 9. **return** the list η_{ehr_clist}
-

Finally, the procedure of searching and retrieving EHRs across states within a country is similar to Algorithm 5, where the retrieval request is sent from a hospital super-peer agent β_{HOS} to its state super-peer agent β_{STA} . Agent β_{STA} then initiates the concurrent searches by broadcasting the request to all other state super-peer agents. Each state super-peer agent γ_s , representing state S , performs a permission checking based on the access control policies stored as SR_{ACP} in the state blockchain. If valid, γ_s executes Algorithm 5 to search and retrieve EHRs of patient p from all cities within state S . The return result η_{ehr_clist} is appended to η_{ehr_slist} if it is not empty. When all state super-peer agents have either responded or timed out, the list η_{ehr_slist} is returned and sent back to β_{HOS} via β_{CIT} . β_{HOS} can then use the links to access and retrieve patient p 's EHRs on behalf of the end user.

6. Case Study

To demonstrate the feasibility and efficiency of our proposed approach, we conducted experiments and evaluated the performance of our hierarchical approach based on the settings and results of each simulation. In our experimental environment, we utilized multiple servers and computers connected under the same network. The specifications of the servers include Intel® Core™ i7-4790k CPU @ 3.60GHz (4 CPU Cores); 16 GB RAM, Windows 10 OS (64-bit, x64-based processor); and 256 SSD Hard Drive. Our experimental environment also had a recorded Internet speed of 600 Mbps.

6.1. Numbers of Published Blocks During a Week

In the first case study, we test our temporary and permanent block approach by conducting simulations to evaluate the need to

use temporary blocks. We simulate and analyze the number of permanent blocks that can typically be created each day of a week based on predefined threshold values. These threshold values are the maximum total size of 2GB for all accumulated temporary blocks and a maximum of 100 new blocks added during a single day. If neither of the thresholds is reached, a permanent block is always created at the end of the day. The number of temporary blocks added during a day is determined by the number of patient visits. We assume that a patient visit always results in the generation of an EHR, which is saved as an HR_{MER} and immediately published to the blockchain as a temporary block. To simplify our experiments, we focus only on HR_{MER} rather than other record types, i.e., HR_{AR} , HR_{UPR} , and HR_{ACP} , because in real-world scenarios, HR_{MER} is the main contributor of block content in hospital blockchains. For the content or nature of the EHRs stored in each HR_{MER} , we use the following experimental settings. Each HR_{MER} includes text-based reports in the size range of [5, 10] KB, while there is also less than a 10% probability of including multimedia files in the size range of [10, 500] MB. Thus, an HR_{MER} must contain text-based report along with possible multimedia files of different sizes. The range of patient visits are based on hospital sizes, where we simulate three different sizes of hospitals. The first type of hospitals has daily patient visits of [10, 100] and is categorized as a small hospital. The second type of hospitals has daily patient visits of [50, 500] and is categorized as a medium hospital. The third type of hospitals has daily patient visits of [100, 1000] and is categorized as a large hospital. Table 1 shows the number of patient visits for each day of a week at each simulated hospital.

Table 1: Numbers of Patient Visits per Day

Hospital Size	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Large	900	700	500	550	700	750	850
Medium	450	300	150	200	300	350	400
Small	100	70	50	55	60	75	90

We now conduct experiments to generate permanent blocks based on the number of patient visits per day. Figure 9 shows the average number of permanent blocks that can be formed at each hospital based on the experimental settings and the numbers of daily patient visits listed in Table 1. As we can see from the figure, even for a large hospital, the number of permanent blocks published per day is limited. The time interval between each addition of permanent blocks can be several hours or even longer, depending on the number of permanent blocks added that day. This indicates a critical need to use temporary blocks to publish data to the blockchain in a timely manner for immediate access without delay. Based on the results in Figure 9, we conclude that medium and small hospitals would benefit the most from our temporary and permanent block approach, as they generate the fewest average numbers of permanent blocks. Figure 10 shows the relationship between the number of permanent blocks formed vs. the number of temporary blocks added during a day at a medium-sized hospital. Since each patient visit results in a new temporary block being created, the number of newly added temporary blocks is equal to the number of new patient visits. According to the simulation results, in medium-sized hospitals, when the number of patient visits increases, the number of new permanent blocks also increases. When the maximum number of

patient visits is reached, i.e., 500, the number of new permanent blocks formed daily is between 4 and 7, which is considered to be very acceptable in terms of spatial efficiency.

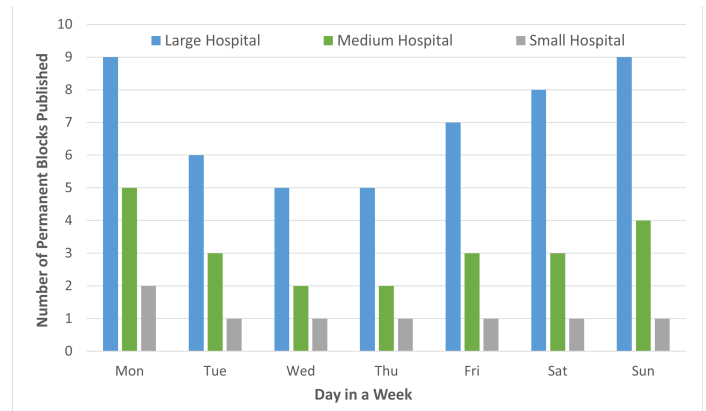


Figure 9: Average Number of Permanent Blocks Published During a Week

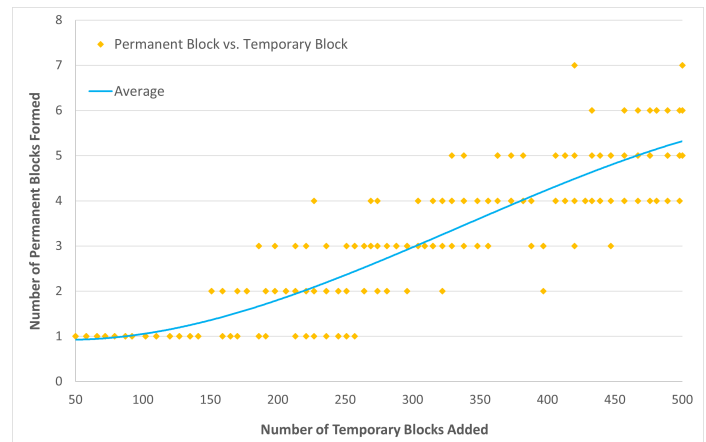


Figure 10: Number of Permanent Blocks Formed vs. Number of Temporary Blocks Added During a Day in a Simulated Medium-Sized Hospital

6.2. Latency of Publishing Temporary and Permanent Blocks

In this experiment, we simulate the creation and addition of permanent and temporary blocks to the blockchain. We record and analyze the time to create, broadcast, and publish temporary and permanent blocks based on a varying number of hospital super-peer agents within an HBN. We use the same settings established in the first case study for this experiment. This means that in a typical day, a new permanent block will have either 100 records stored or 2 GB in size, according to the given thresholds. Each temporary block includes only HR_{MER} containing text-based reports and potentially (10%) multimedia files of [10, 500] MB in size. In addition, we add random delays in the range of [100, 3000] milliseconds to simulate network congestion during the consensus process. Figure 11 shows the experimental results and the efficiency of our approach using temporary blocks. From the figure, we can see that at most, it takes about less than half a minute to create a temporary block and add it to the blockchain. Due to the large variation in the potential size and frequency of temporary blocks containing multimedia files in our experimental settings, the range between each case can vary considerably. In contrast, for permanent blocks, the range is more consistent for each case due to previously determined size and number thresholds. In general, the overall time for a permanent block to

be created and added to the blockchain is less than one minute. While there is no big data transferred during the consensus process, additional validation is required by the other super-peer agents to verify the permanent block broadcast by the block announcer and the associated temporary blocks previously stored in their local copies of the blockchain. This significantly increases the overall time required for the consensus process, which takes longer time when compared to the publication process involving temporary blocks only. Nevertheless, it takes no more than 20 seconds to create and add a temporary block to the blockchain, which allows many consensus processes for new temporary blocks to be performed during a single day without encountering significant delays. Therefore, we can conclude that our approach supports efficient creation and addition of temporary and permanent blocks to the blockchain.

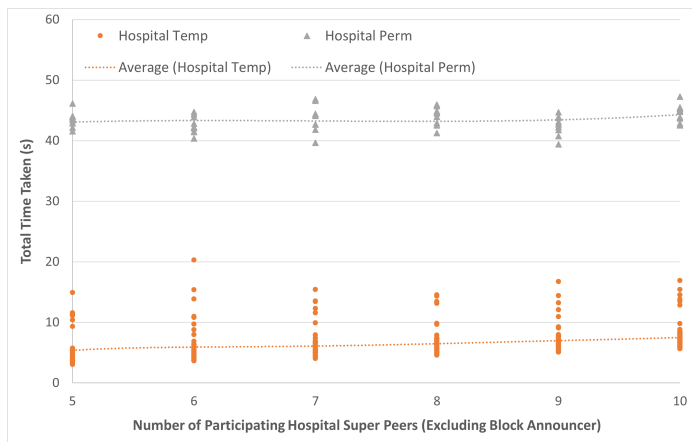


Figure 11: Total Time Taken to Create, Broadcast, and Publish Hospital Temporary and Permanent Blocks

6.3. Search and Retrieval Time in Hierarchical Blockchains

In this experiment, we simulate the integrated concurrent search and retrieval processes for patient EHRs. We record and analyze the total time taken to search and retrieve a number of patient’s EHRs in the hierarchical blockchain networks. Our experimental environment consists of three layers of fully simulated hierarchical networks with a varying number of HBNs, CBNs, and an SBN. We also have a range of [6, 10] hospital super peers within an HBN, [100, 500] city super peers within a CBN, and 50 state super peers within the SBN. The contents of our hospital, city and state blockchains contain all the necessary or relevant access control policy records to enable our searching process. To simplify the overall downloading process, the hospital blockchain contains only EHRs with multimedia files. Each EHR has a range of [10, 500] MB in size, similar to our previous case studies. The EHRs are stored in different hospital blockchains within different HBNs to simulate a patient who visits multiple hospitals in different cities. We also introduce a random delay with a range of [100, 3000] milliseconds to simulate network congestions. In addition, we assume that all hospitals have the required infrastructure to allow multiple concurrent file downloads to mitigate throttling when any number of peers download multiple files simultaneously. Each hospital agent also maintains a separate local index file for efficient responses to any EHR-related inquiries. Figure 12 shows the total time taken to search and retrieve different numbers of EHRs. Based on the

figure, we can see that the search time remains relatively constant regardless of the number of EHRs to be searched. Several factors, such as the use of separate index files to track patient EHRs for fast response and the small size of the metadata involved during the concurrent search process, contribute significantly to this stability. However, when the total numbers of EHRs to be retrieved increases, the time to retrieve those EHRs also increases. This result is consistent with the experimental results we reported in our previous work [7]. However, the current approach is more efficient compared to the previous method as the waiting time for creating new access policy records is not needed any more after the search process. This leads to an overall time improvement, which allows multiple EHRs to be searched and retrieved in one integrated process.

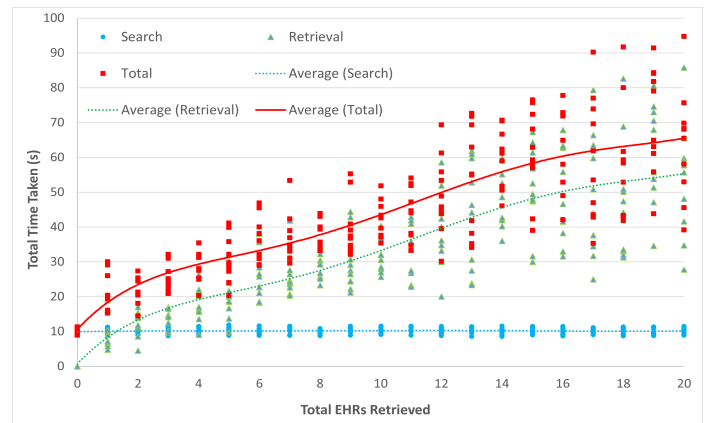


Figure 12: Time Taken to Search and Retrieve Varying Number of EHRs

7. Conclusions and Future Work

In this paper, we build on the concepts and methods of previous work [7] and further explore them by introducing several design changes. These changes include limiting the number of hospital super-peer agents in an HBN, adopting a temporary and permanent block scheme, and integrating the search and retrieval processes for EHRs. The number of hospital super-peer agents in an HBN is limited to minimize the redundancy of big data stored in the blockchain. This also ensures better scalability, as we limit the growth potential of hospital blockchain to a more manageable level. The use of the temporary and permanent block scheme in our hierarchical blockchain approach ensures timely publications of EHRs in an HBN. Any urgent data can be published in a temporary block immediately, while once a certain threshold has been reached, a permanent block consisting of a number of temporary blocks can be formed. This scheme allows for timely and space-saving publications of EHRs to the cloud-based hospital blockchains. Finally, the search and retrieval processes for EHRs have been integrated to be more efficient. As the experimental results show, our new hierarchical blockchain approach is efficient and effective, allowing for a timely and spatially efficient publication of EHRs to the hospital blockchain as well as a better overall performance in the integrated search and retrieval process of EHRs.

In future work, we plan to perform an in-depth comparison of our cloud-based on-chain blockchain approach with IPFS-based off-chain approaches [10], [11] and mechanisms for reliable and secure distributed cloud data storage [23]. We will focus on the

redundancy and efficiency aspects of such comparisons and evaluate the performance of our cloud-based hierarchical blockchain mechanism. In addition, we plan to further improve and develop our approach to defend against real-world attacks, such as DDOS attacks and insider threats [24], [25], [26]. An emphasis will be on evaluating the performance of our consensus process and cryptographic procedures against potential attacks and improving them if necessary.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] S. Nakamoto, "Bitcoin: a peer-to-peer electronic cash system," October 2008. Retrieved on January 15, 2021 from <https://bitcoin.org/bitcoin.pdf>.
- [2] O. Dib, K.-L. Brousmiche, A. Durand, E. Thea, E. B. Hamida, "Consortium blockchains: overview, applications and challenges," *International Journal on Advances in Telecommunications*, **11**(1&2), 51-64, 2018.
- [3] M. T. de Oliveira, L. H. A. Reis, R. C. Carrano, F. L. Seixas, D. C. M. Saade, C. V. Albuquerque, N. C. Fernandes, S. D. Olabarriaga, D. S. V. Medeiros, D. M. F. Mattos, "Towards a blockchain-based secure electronic medical record for healthcare applications," in *Proceedings of the 2019 IEEE International Conference on Communications (ICC)*, 1-6, Shanghai, China, May 2019, doi: 10.1109/ICC.2019.8761307.
- [4] Q. Xia, E. B. Sifah, K. O. Asamoah, J. Gao, X. Du, M. Guizani, "MeDShare: trust-less medical data sharing among cloud service providers via blockchain," *IEEE Access*, **5**, 14757-14767, July 2017, doi: 10.1109/ACCESS.2017.2730843.
- [5] A. Thamrin, H. Xu, "Cloud-based blockchains for secure and reliable big data storage service in healthcare systems," in *Proceedings of the 15th IEEE International Conference on Service-Oriented System Engineering (IEEE SOSE 2021)*, 81-89, Oxford Brookes University, UK, August 2021, doi: 10.1109/SOSE52839.2021.00015.
- [6] R. Ming, H. Xu, "Timely publication of transaction records in a private blockchain," *2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, 116-123, Macau, China, December 2020, doi: 10.1109/QRS-C51114.2020.00030.
- [7] A. Thamrin, H. Xu, "Hierarchical cloud-based consortium blockchains for healthcare data storage," in *2021 IEEE 21st International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, 644-651, Hainan Island, China, December 2021, doi: 10.1109/QRS-C55045.2021.00098.
- [8] S. Alexaki, G. Alexandris, V. Katos, N. E. Petroulakis, "Blockchain-based electronic patient records for regulated circular healthcare jurisdictions," in *Proceedings of the 23rd IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 1-6, 2018, doi: 10.1109/CAMAD.2018.8514954.
- [9] R. Kumar, N. Marchang, R. Tripathi, "Distributed off-chain storage of patient diagnostic reports in healthcare system using IPFS and blockchain," in *Proceedings of the 2020 International Conference on COMMunication Systems & NETWORKS (COMSNETS)*, 1-5, Bengaluru, India, 2020, doi: 10.1109/COMSNETS48256.2020.9027313.
- [10] D. Li, W. E. Wong, M. Zhao, Q. Hou, "Secure storage and access for task-scheduling schemes on consortium blockchain and interplanetary file system," *2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, 153-159, 2020, doi: 10.1109/QRS-C51114.2020.00035.
- [11] Y. Jeong, D. Hwang, K. Kim, "Blockchain-based management of video surveillance systems," in *Proceedings of the 2019 International Conference on Information Networking (ICOIN)*, 465-468, Kuala Lumpur, Malaysia, 2019, doi: 10.1109/ICOIN.2019.8718126.
- [12] Z. Su, H. Wang, H. Wang, X. Shi, "A financial data security sharing solution based on blockchain technology and proxy re-encryption technology," in *Proceedings of the IEEE 3rd International Conference of Safe Production and Informatization (IICSPI)*, 462-465, Chongqing City, China, 2020, doi: 10.1109/IICSPI51290.2020.9332363.
- [13] H. Wang, Y. Song, "Secure cloud-based EHR system using attribute-based cryptosystem and blockchain," *Journal of Medical Systems*, **42**(152), 1-9, July 2018, doi: 10.1007/s10916-018-0994-6.
- [14] B. S. Egala, A. K. Pradhan, V. R. Badarla, S. P. Mohanty, "Fortified-chain: a blockchain based framework for security and privacy assured Internet of medical things with effective access control," *IEEE Internet of Things Journal*, **8**(14), 11717-11731, July 2021, doi: 10.1109/JIOT.2021.3058946.
- [15] A. Fernandes, V. Rocha, A. F. d. Conceicao, F. Horita, "Scalable architecture for sharing EHR using the Hyperledger blockchain," in *Proceedings of the IEEE International Conference on Software Architecture Companion (ICSA-C)*, 130-138, Salvador, Brazil, March 2020, doi: 10.1109/ICSA-C50368.2020.00032.
- [16] N. Nicol, H. Xu, "A blockchainless approach for trusted public construction bidding," *Computer and Information Science Technical Report*, Computer and Information Science Department, University of Massachusetts Dartmouth, December 2018.
- [17] L. Cui, S. Yang, Z. Chen, Y. Pan, M. Xu, K. Xu, "An efficient and compacted DAG-based blockchain protocol for industrial Internet of things," *IEEE Transactions on Industrial Informatics*, **16**(6), 4134-4145, 2020, doi: 10.1109/TII.2019.2931157.
- [18] A. Buzachis, A. Celesti, M. Fazio, M. Villari, "On the design of a blockchain-as-a-service-based health information exchange (BaaS-HIE) system for patient monitoring," in *Proceedings of the IEEE Symposium on Computers and Communications (ISCC)*, 1-6, Barcelona, Spain, July 2019, doi: 10.1109/ISCC47284.2019.8969718.
- [19] D. C. Nguyen, P. N. Pathirana, M. Ding, A. Seneviratne, "Blockchain for secure EHRs sharing of mobile cloud based e-health systems," *IEEE Access*, **7**, 66792-66806, May 2019, doi: 10.1109/ACCESS.2019.2917555.
- [20] H. Guo, W. Li, M. Nejad, C. Shen, "Access control for electronic health records with hybrid blockchain-edge architecture," in *Proceedings of the 2019 IEEE International Conference on Blockchain (Blockchain)*, 44-51, July 2019, doi: 10.1109/Blockchain.2019.00015.
- [21] H. Guo, W. Li, E. Meamari, C. Shen, M. Nejad, "Attribute-based multi-signature and encryption for EHR management: a blockchain-based solution," in *Proceedings of the 2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 1-5, Toronto, ON, Canada, May 2020, doi: 10.1109/ICBC48266.2020.9169395.
- [22] M. Meingast, T. Roosta, S. Sastry, "Security and privacy issues with health care information technology," in *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society*, 5453-5458, New York, NY, USA, September 2006, doi: 10.1109/IEMBS.2006.260060.
- [23] H. Xu, D. Bhalerao, "Reliable and secure distributed cloud data storage using Reed-Solomon codes," *International Journal of Software Engineering and Knowledge Engineering (IJSKE)*, **25**(9&10), 1611-1632, 2015, doi: 10.1142/S0218194015400355.
- [24] S. Northcutt, J. Novak, *Network intrusion detection*, 3rd Edition, Sams Publishing, August 2002.
- [25] H. Xu, A. Reddyreddy, D. F. Fitch, "Defending against XML-based attacks using state-based XML firewall," *Journal of Computers (JCP)*, **6**(11), 2395-2407, November 2011, doi: 10.4304/jcp.6.11.2395-2407.
- [26] J. Mirkovic, P. Reiher, "A taxonomy of DDos attack and DDos defense mechanisms," **34**(2), 39-53, April 2004, doi: 10.1145/997150.997156.

Towards a Framework for Organizational Transformation through Strategic Design Implementation

Lynne Whelan^{1,*}, Louise Kiernan², Kellie Morrissey², Niall Deloughry²

¹South East Technological University Ireland, Carlow, R93V960, Ireland

²Univeristy of Limerick, School of Design, V94T9PX, Ireland

ARTICLE INFO

Article history:

Received: 28 January, 2022

Accepted: 07 April, 2022

Online: 22 April, 2022

Keywords:

*Transformational
Design
Strategy*

ABSTRACT

The aim of the research is to contribute to the emergent field of strategic design as an approach to creating transformational impacts in organizations. This is driven by international strategies to promote sustainable business environments which are innovative and adaptive to change. The literature identifies a gap in the knowledge in relation to how knowledge flows within an organization from knowledge producing activities and back. This flow between management and operational activity is identified as the ongoing innovation capability and ultimately the area of transformational impact within the organization. However, the structures which support this flow and how it is measured are ambiguous and ill defined. Strategic design is a holistic approach to developing business strategies which may provide methods which support the flow of knowledge and provide the transformational impacts. A 'research through design' approach was taken to collect data from small to medium enterprises engaging in a series of strategic design workshops. Analysis was carried out through visual mapping of how strategic design is applied in an organization at management and operational level. This identifies the nuanced differences in application and design tools used to develop management strategy versus operational strategy. It also identifies the points of knowledge creation and transfer which builds business intelligence to inform both business and operational strategies. This is presented as a contextual framework which provides the basis for understanding the complexity of the flow between the two. This contributes to informing organizations of where the links between these strategies may be built, resulting in a more dynamic organization which is nimble, innovative, and adaptive to change.

1. Introduction

This research is an extension of work originally presented in 2021 IEEE Technology & Engineering Management Conference. The title of the previous paper was 'Measuring the Success Factors of Strategic Design Implementation'[1]. This paper presents a literature review of discourse on the development and application of strategic design across the business landscape with a focus on transformational impacts. The gap in the knowledge is identified in relation to how the flow of organizational knowledge which is both tacit and explicit is supported and leveraged to create strategic growth and innovation. A 'research through design' approach is used for data collection and analysis through a series of design workshops. The findings are presented in a contextual framework of where and how strategy is used in an organization and where

and how a design approach supports the transformational flow of knowledge. The research began with seeking to understand methods of determining success factors of strategic design implementation. A design approach is a holistic view of a situation or environment. It looks at the usability, feasibility, and viability factors to solve problems [2]. This can be a people centered approach, a technological approach, and a business approach. Therefore, to consider best practices for measuring success factors in a way that reflects this, different disciplines were considered in the literature review such as social science, data science and business. The literature review noted that across these disciplines they all referred to a flow of knowledge between management strategy and operational strategy. The gap in the knowledge is in relation to how the flow of knowledge operates within an organization. Knowledge management is described as the process of acquisition, storage, distribution and the use of knowledge [3]. The research design was to collect data from seven small to

*Corresponding Author: Lynne Whelan, Lynne.Whelan@itcarlow.ie

medium enterprises (SMEs) in research through design approach as they engage in a series of strategic design workshops within a Technology Gateway in an Irish University. The data is analysed through visual mapping to provide insights into the nuanced differences in approach to business strategy and operational strategy and the relevance of knowledge exchange between the two. The visual mapping highlights the tools used across a four-stage process of strategic design application. The findings highlight the process as an ongoing cycle of innovation and identifies the key points of knowledge exchange. It is from this that the author has developed a framework of organizational strategy depicting the strategic design process, tools used and knowledge acquisition and transfer points. The research aims to inform both the design and business communities nationally and internationally, in the actionable methods, tools, and processes for accessing and implementing a strategic design approach for transformational impacts.

2. Literature Review

There are clear international drivers to promote innovation on economic, social, cultural and educational sectors. Innovation is recognized as a means to create sustainable organizations which can not only problem solve to adapt to change but also become the drivers of change. The US Chamber of Commerce Global Innovation Policy Centre (GIPC) champions innovation to “create jobs, save lives, advance global economic and cultural prosperity, and generate breakthrough solutions to global challenges” [4]. This highlights the significant and varied applications for innovation. The Organization for Economic Co-operation and Development (OECD) work to improve innovation across 38-member countries to stimulate economic progress and world trade. The core mission of OECD is not static but needs to respond to an evolving and challenging world and notes that this is a time when the expanding integration and influence of new technologies are disrupting both advanced and emerging economies [5]. It is highlighted that adaptability to change, and innovation are at the core of global economic strategies to develop competitive and sustainable organizations. This paper is set within the context of Irish Higher Education (HEI) engagement with industry which has long been identified as a model to foster and promote research and development, innovation and faster knowledge exchange between researchers and industry. The publication of ‘Winning by Design’ marks the clear linking of design and Innovation by the Irish State [6]. Design was recognized as a process of innovation and design thinking as a strategic tool for innovative business development. It is highlighted that it will require more than transactional interventions and that the focus should be on longer term developmental programmes to achieve transformational impacts.

“Mobilizing universities needs to be addressed in a holistic way and not just by focusing on transactional interventions such as consultancy services for local companies. It is tempting to focus on transactional mechanisms as they have clear outputs such as number of firms assisted. However, they are less likely to have the longer-term outcomes and impacts that can be achieved with ‘transformational’ and more developmental programmes” [7]

In seeking to understand transformational engagements within a design context, strategic design processes and approaches were

examined through literature. Roberto Verganti presents a theory of design driven innovation as radically innovating what things mean by bringing the designer into the corporate space [8]. This approach however is referenced in the context of new product development within the business model. Giulio Calabretta who writes specifically about strategic design describes it as “the professional field in which designers use their principles, tools and methods to influence strategic decision making within an organization” [9]. However, Calabretta also leans towards the strategic input as an underpinning of the product brief “designers are no longer mere executors of design briefs-they are involved in the crafting of the briefs and guide the strategic decisions that underlie them”. This research paper argues that the impacts may be further reaching and more transformational than product strategy alone. It presents strategic design as a holistic approach which when applied to industry can result in not only new products but new services, new markets, or simply new ways of doing things, impacting an organization across business and operational strategy. Strategic design is the application of design thinking to develop strategies, in a business setting this is primarily strategies for growth and innovation. An approach to better understanding the application of design on both product development and business model strategy is to consider the outcomes and the measures of success factors. In other words what are the outcomes of a strategic designer’s engagement when applied to new product development or when applied to the broader business model strategy. A new product is an explicit and tangible outcome of a design engagement but how do we measure the success factors of the business model strategy. This area is more difficult to communicate and measure in terms of impacts as it may be of a tacit nature, less tangible and more subjective. It would appear that a more qualitative method of assessing the outcomes may be of value.

2.1. Social science, data science and business

Design approaches encompass both tacit and explicit knowledge to form holistic understanding. Whilst explicit knowledge is objective and easy to quantify, tacit knowledge is subjective and difficult to quantify and therefore often overlooked. Literature was therefore considered from disciplines which measure values in different ways, such as social science, data science and business. In social science, Etienne Wenger was one of the first to use the term ‘communities of practice’ – groups of people who together accumulate and share their collective learning [10]. In relating to organizations, Wenger considers ways to measure the value of the organizations community and proposes that good measurement has to follow the course of the story. He refers to the process of analysing the stories as “systematic anecdotal evidence” and that by considering communities of practice, organizations have the chance to see the value in qualitative not quantitative terms. Wenger concludes “in fact , the value of knowledge is a flow from knowledge producing activity to performance and back”. Wenger specifically links social practice and behaviours with a transformative business model. He refers to the generation of social practice and changing the designs of our organizations so that they are more in line with our behaviours. Geoff Walsham also looks at information flow within an organization but considers it from a context of information systems such as data processing. “in order for all types of

organizations to succeed, they need to be able to process data and use information effectively” [11]. Walsham refers to this information as informing everyday operations such as planning, controlling, organizing and decision making. This process of information gathering and application, whilst dealing with data science, is also reflective of Wenger’s communities of practice and the information flow between operations and business strategy. As far back as the 1970s, Mintzberg was one of the first to highlight the ‘flaw’ in organizational structure which is how work, information and decision processes actually flow through it [12]. In other words, the interconnectedness and flow between business strategy and operations. Wenger follows this understanding and links the communities of practice to managing the knowledge in an organization, between employees and external stakeholders as peer exchange. He is clear however that to nurture knowledge resources a CEO must understand the broad strategy but also needs to be in contact with the practitioners who manage that knowledge. This infers a more involved and human influence on building the strategy. In exploring these elements, we can consider praxeology, the theory of human action and purposeful behaviour, and axiology, the philosophical study of value. Dr Marina Pankina, presents a paper specifically looking at axiology and praxeology of design thinking. Pankina states that “design combines an objective and subjective approach. This is what distances it from the classical science that aims for objectivity and elimination of everything subjective, as a key to the validity of knowledge” [13]. This is particularly relevant in a design context as design is focused on understanding cultural, ethical, and economic values, along with human drivers, motivations, wants and needs. It may be that these human influences are the tacit links to successful transformative strategies. Additionally, this would indicate that to measure the success factors of design engagements it is also appropriate to combine objective and subjective results. Reflecting on this literature we can see there are, according to Walsham and Wenger, two key tiers of approach to organizational strategy; a managements’ strategic level of broad initiative and an operational level of knowledge creation and that the flow between the two is essential (Figure 1). Measuring the success factors can be approached subjectively based on systematic anecdotal evidence or objectively based on data collection, application, and management. However, Wenger clearly links social practice and behaviours with transformative business models. Pankina presents design as a methodology which combines both an objective and subjective approach and that designers work to understand the human factors of behaviour and value to bring meaning.

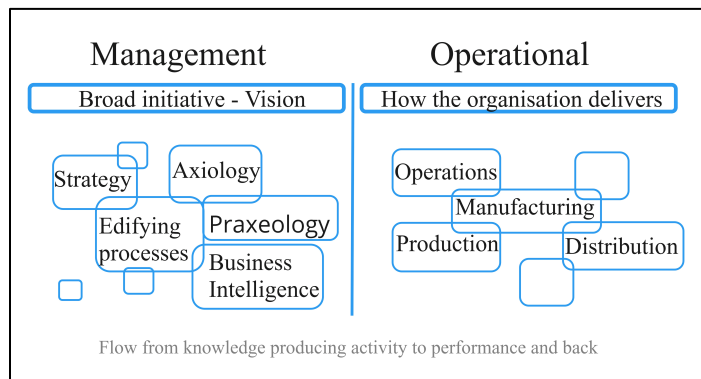


Figure 1. Two tiers of approach for organizational strategy (authors own 2021)

3. Research Question

This combined objective and subjective approach that design offers, applied holistically across an organization at management and operational levels, may result in transformational impacts. The research question is; How can the strategic design approach support the flow of knowledge (both tacit and explicit) across management and operational strategies resulting in transformational impacts?

4. Methodology

The research is approached in a ‘research through design’ methodology. This is a research approach that uses design practice to inform research [14]. This research was carried out in the Design + Technology Gateway which specializes in design research, design driven innovation and strategic design. The research design was to participate in the facilitation of the design strategy workshops with a series of small to medium enterprises (SMEs). Each SME owner/manager engaged in four workshops facilitated by two design strategists as laid out in Table 1.

Table 1. Research design

Location	Design Studio & remote video workshops
Participants	Design strategists x2 SME manager x1
Time	2 hours x 4 weeks
Duration	6-8 weeks per SME (x7 SMEs)
Methods	Visual drawing/text on white wall, post-it notes, colored pens. Use of design toolkit
Total Organization	7
Sectoral range	Corporate service, Brewing and distilling, IOT, Health care, Retail, Events, Technology sector

The workshops were carried out in the design studio with large whiteboard walls which were used to capture the data as the participants engaged in the process. Colored markers, post it notes, print outs and digital images were used as a mixed media approach to capture data (Figure 4). Strategic design, as delivered by the Design+ team consists of a four-stage process (Figure 3). This process is derived from the underlying theoretical double diamond design process [2]. This is an approach which is internationally recognised by both industry and academic realms and is based on a series of divergent and convergent thinking (Figure 2).

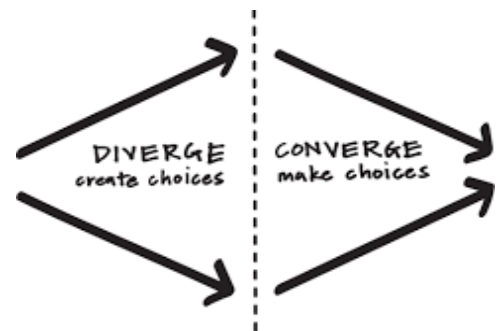


Figure 2. Divergent and convergent thinking [IDEO 2019]

The theoretical approach of convergent and divergent thinking has been translated by the design strategy team into the business landscape as an actionable four stage process for the development of innovation strategies.

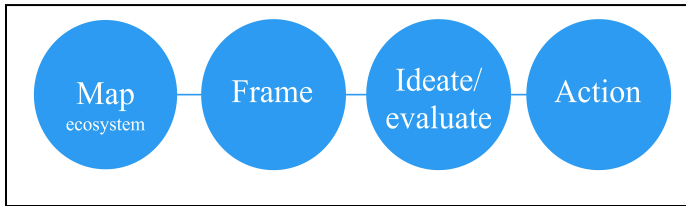


Figure 3. Design+ Technology Gateway, four stage strategic design process (authors own 2020)

The process is supported by a design strategy toolkit. The toolkit combines business tools, design tools and tools newly developed by the team which responded to an identified business need. Examples of these are user journey maps, role reviews, Ansoff matrix, leadership wheel, strategic roadmap. The tools are generally A3 infographic prompts to ensure the thinking is being challenged at each step of the way and that all things are being considered (Figure 5). The tools differ from templates in that they are flexible and can be used in different ways unlike a template which is predefined.



Figure 4. Mixed media approach to data capture (authors own 2021)



Figure 5. Strategic design toolkits (authors own 2021)

4.1. Data collection

The research began with three SMEs and was specifically looking at ways to measure the success factors of a strategic design engagement. The objective was to analyze the outcomes of

the engagement which were captured in a strategic roadmap, to better understand the resulting values (Figure 5).

Table 2. SMEs 1-3 Background

<p>SME1 was as a legacy business on the point of handover to the next generation which incorporated three retail outlets. A new strategy was required based on a new vision by the next generation, considering challenges within today's markets but also with a consideration for future proofing and creating a sustainable business model fit for purpose. As part of the future vision and in developing options for a new way of doing things, a specific scanning technology was identified around which there were many uncertainties which required evaluation.</p>
<p>SME2 was a start-up business with a specific offering in the technology sector that was not clearly linked to a specific target market. There was therefore no clear understanding of individual user needs. Their request on engaging with the team was to frame key market sectors and develop a strategy for responding to their needs.</p>
<p>SME3 was an established events-based company who was seeking a new growth strategy. The company required support with ideation and rationale building around diversification. They had also identified an opportunity to leverage the existing company data to improve business intelligence.</p>

The research captured data from each stage of the process through visual mapping on whiteboard walls and digital whiteboards. The background context to the engagement requirements was recorded in Table 2. The key area of data capture was in the culminating strategic roadmaps for each of the SMEs. The strategic roadmaps present the innovation or the new strategy through steps to implementation, resources required, participation scope, timeframes, key decision points and milestones. This provided the research with an insight into a variety of outcomes across the SMEs such as collaborate with new partner, run a pilot study, engage with data analytics, or apply for funding support. In addition, there were noted new approaches to strategy such as the introduction of new strategy meetings within the organization, who would be involved and what tools could be used to support those meetings. The research also recorded any identified new ways of doing things such as introduction of digital whiteboards and visual mapping techniques as an approach to dealing with complex data. Also considered was the introduction of new platforms for creating shared context and knowledge exchange utilizing tools from the toolkit. Material outputs were also recorded such as the digital whiteboards, infographics, the toolkit, the roadmap, and a presentation of the process. In building on the first three SMEs phase of research an additional four SMEs were engaged in a series of strategic design workshops. This brings a total of seven SMEs and a cumulative time of fifty-six hours of workshops from which data was collected. The objective of this additional research was to begin assessing the nuanced differences in approach to applying strategic design to a broad business strategy and to an operational level strategy. The data collected focused on the tools used when applying strategic design to both the broad business model and on an operational level. The tools represent the area of focus pertinent to the objective and were based on the Design+ strategic design toolkit which includes almost 40 tools.

4.2. Analysis

Analysis of the first three SMEs data collected was carried out through visual mapping. Visual mapping enables grouping of commonalities and differences and themes to emerge as visual cues. The workshops act as a process of co design therefore the roadmaps are developed and validated with the business owners. The actions identified in the roadmaps were listed and grouped into tacit and explicit responses and then further categorized into physical material outputs, tangible outcomes, and potential tacit impacts as presented in Figure 6. It was also possible to consider the transactional elements and the transformational elements of the resulting outcomes. The transactional elements are those which are task related and focused on planning and execution, as opposed to the transformational elements which influence change of the existing culture of the organization.

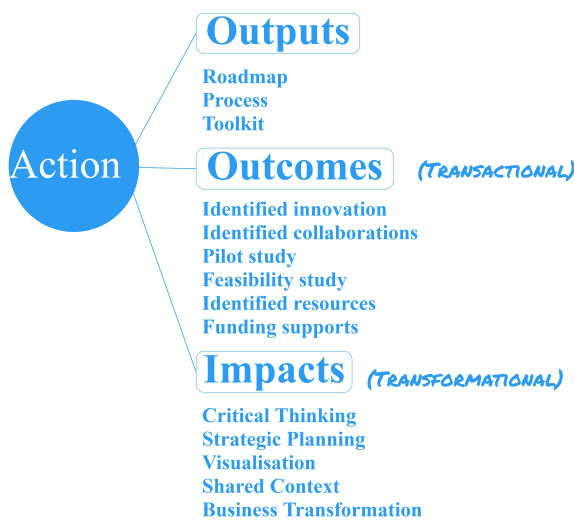


Figure 6. Analysis of roadmap conclusions as outputs, outcomes, and impacts (authors own 2021)

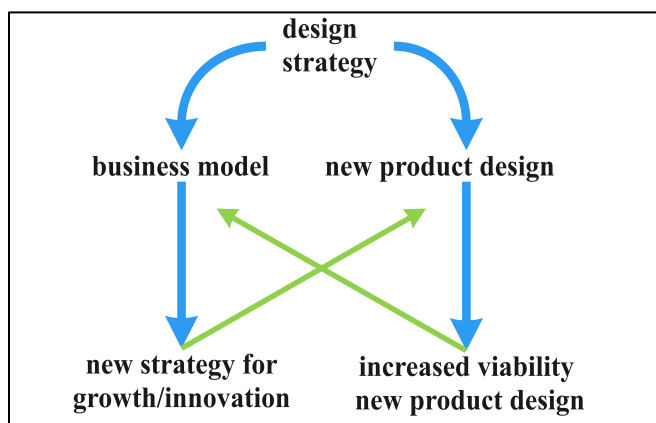


Figure 7. Garden gate model of flow between business and product strategies (authors own 2021)

The analysis of the additional four SMEs which was the extended research, was also carried out through visual mapping of all of the tools used throughout the process with the aim of detecting the nuanced differences in approach to business level and operational strategy development (Figure 9). Each of the tools

highlighted in the state of play and action stages represent the questioning and data gathering activity. As noted in the literature review, strategic design is commonly associated with new product development. Considering the models from this context, and analyzing both approaches, it became apparent that the business strategy may result in a new product concept which then must undertake the operational strategy route to develop the product. Similarly, if a new product concept is developed, the new business model must go through the business strategy development, represented in the ‘Garden Gate’ model (Figure 7).

This is a potentially ongoing cycle of adaptability and innovation which reflects the flow between business and operations. By considering this in the broader context of business and operations, a model was created, the triple transverse model, (Figure 8) which provides a contextual representation of organizational strategy. In building on this, the activities that happen when applying the strategic design process across business and operational strategies were mapped. The author developed this with the addition of the knowledge creation and transfer points that occur between business and operations to produce the triple transverse framework (Figure 9). This is a key framework for the research and for our understanding of transformational impacts. It provides a contextual reference of how strategic design is applied and where the knowledge is acquired, distributed, and leveraged for ongoing adaptability to change and innovation within the organization.



Figure 8. Triple transverse model of organizational strategy (authors own 2021)

5. Findings

The aim of the research is to contribute to the emergent field of strategic design as an approach to creating transformational impacts in organizations. The findings show that firstly, there are nuanced differences in the application of strategic design to develop business strategy or to develop an operational strategy. Capturing this through visual mapping and as the triple transverse framework will inform designers and managers of the approach and the specific tools to ensure all opportunities are presented when innovating. Secondly, innovation by its nature has unknown outcomes, having an indicative guide to outputs, potential outcomes and typical impacts will assist in communicating the type of results to expect from a strategic design engagement. Thirdly, there are specific points in the process which, if the appropriate structures and supports are in place, will produce an ongoing cycle of innovation and adaptability, a transformational impact for an organization.

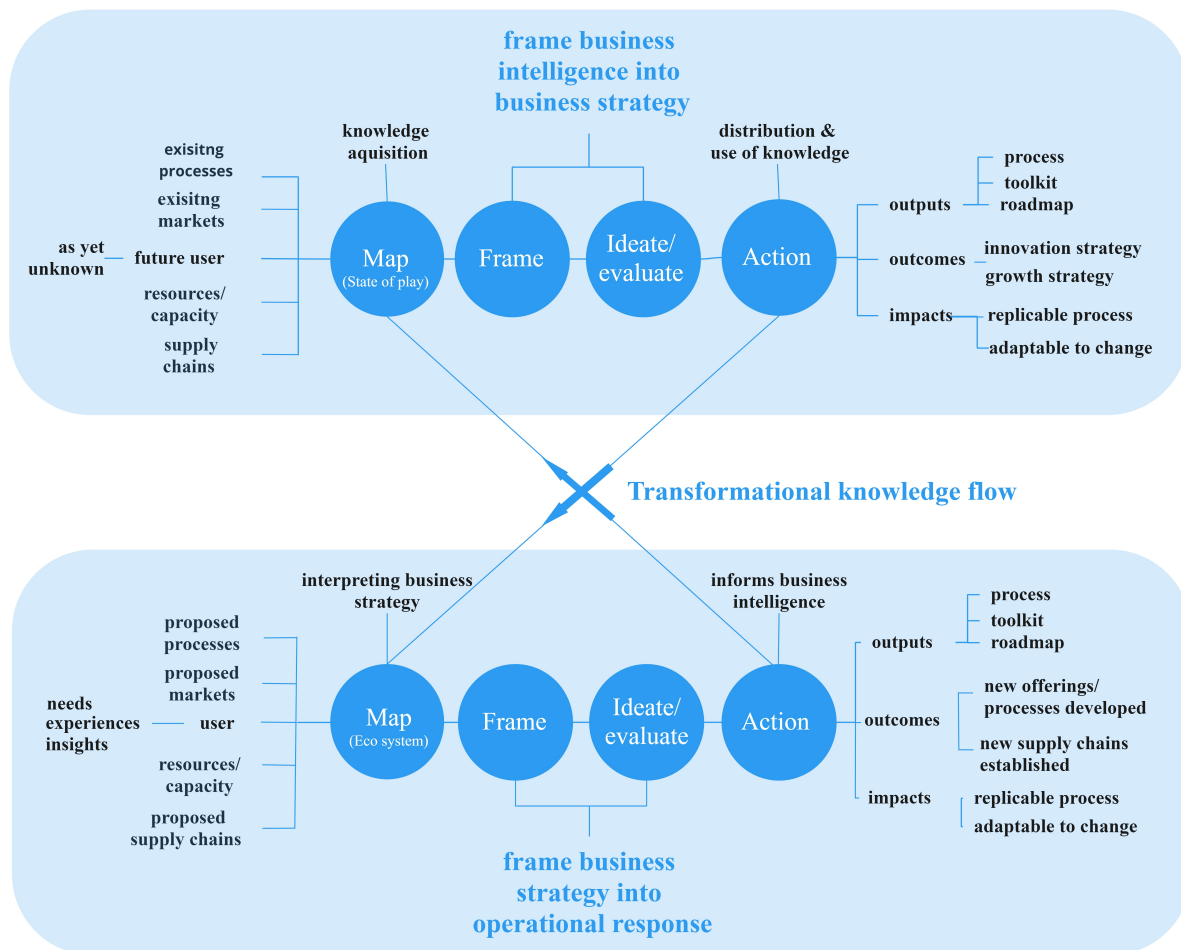


Figure 9. Triple transverse framework of strategic design applied to organizational strategy (authors own 2022)

The categorization of the results into outputs, outcomes and impacts was based on the physical material results classed as outputs, the tangible new direction as outcomes, and the tacit elements as impacts. Throughout the process SME1 framed the human factors and needs to assist in identifying the appropriate scanning technology. They were then introduced to the technical and financial resources within the research network to pursue a feasibility study which included a pilot of the technology. They also formulated a business strategy based on a framed vision for innovation and adaptability. The company introduced strategic meetings between management teams and identified the specific tools to use which could provide a *neutral* voice within family business management. SME2 identified a new application of a technology which was the most suitable in aligning with owner/managers skill sets. As part of the evaluation around viability, a growth plan was formulated and the steps to implementation added to the strategic roadmap. The company vision was framed and the strategy to achieve it was captured. A collaboration was initiated within in the newly identified market sector bringing together the technology and finance sector for a unique new offering. SME3 mapped out the user and business needs for improved interaction and engagement. A CRM system could then be identified appropriate to those needs. Improved business intelligence and decision making ability was framed into usable dashboards for the new system. The company introduced

strategy sessions to continue working towards their operational goals with the use of the toolkit. All three SMEs who participated shared similar material outputs from the engagement such as; a four-stage process, toolkit, infographics, and roadmap. The outcomes varied pertaining to the nature of the innovation, but included results such as; an identified innovation, pilot study, feasibility study, identified collaboration, or new process introduced. The impacts are more difficult to communicate as they are of a tacit nature. However, based on the data collection and analysis we can gain insights to key elements such as: a developed understanding of human drivers and aspirations, the introduction of critical thinking, the use of visualization techniques, and the development of shared context.

In building on this, the triple transverse framework highlights that to measure the success factors of the transformational elements of strategic design engagements, we need to measure the structures within the organization that engage in strategy. The three approaches to strategy are identified as business strategy, operational strategy, and transformational strategy. The transformational strategy depends on the flow of knowledge exchange between management and operations. The research aimed to map the strategic design approach for business strategy development and operational strategy development. The objective was to identify the nuances of implementation from each other

and from traditional business modelling. The findings from mapping the two processes across four SMEs, highlight that within a design strategy approach to operations, the eco system is mapped around the end user needs. This is to produce insights that can be translated into a meaningful product/service response. In a design strategy approach to business strategy development, the future user is unknown. If the process is restricted to existing users, it will limit the opportunities for transformational big system changes which may result in new ways of doing things, new market arenas or new product arenas to develop into. Therefore, the business eco system is mapped to identify areas of opportunity. The mapping includes the interplay between the business model, the existing user groups, the broad stakeholder needs, environments, processes, and experiences. This is a key insight to the nuanced difference in approaches. However, a cycle of interplay between business and operations is apparent. The strategic design approach to developing a business strategy reaches an 'action' stage, and a roadmap is produced. The roadmap provides the steps of implementation which forms the basis to develop the operational strategy. Equally when the operational strategy is developed and reaches the 'action' stage, the information and experiences of implementation are captured in the 'state of play' tools which informs the ongoing business strategy development. This is the ongoing cycle of knowledge within the organisation. The tools used by strategic designers support this process in a holistic way ensuring the capture of both tacit and explicit knowledge and the use of visual prompts and methods to easily exchange information and support the flow of knowledge. Critical thinking, for example, is supported using multiple dialectic tools which look at different perspectives of the same problem. Developing a shared context is supported with tools such as the roadmap tool which provides visual representation of complex data. The roadmap will mark the steps to implementation, participation scope, timelines, milestones, key decision points, resources required etc. This provides a shared context for all stakeholders when undertaking a new direction. It also enables quick changes to be made as everyone can see where the change is required and why and who it may impact. This cycle of innovation between business and operational strategy provides an adaptability and sustainability of a business model which is innovative, competitive, and responsive to change. This is the ongoing cycle of knowledge within the organization.

6. Conclusion

International strategies strive to promote sustainable business practices which are innovative and adaptive to change. The literature review highlights that understanding and leveraging the flow of knowledge within an organisation has a transformational impact. The gap in the knowledge is in understanding the structures that support this transformational dynamic. The research captured and analysed the specific nuances in the application of strategic design as an approach to developing innovation strategies, applied at both management and operational level. The findings have been developed into a triple transverse framework which demonstrates the application of the process and tools at both management and operational level. The framework also highlights where knowledge is acquired and disseminated forming an ongoing cycle of innovation and adaptability to change. This framework can be utilised in

organisations globally, with design facilitation, across all sectors such as business, policy, education, and social applications. The framework contributes to the emergent strategic design sector providing a contextual reference for the application of strategic design for transformational impacts across an organization. The framework is currently being developed into a strategic design for innovation accredited training programme with an accompanying toolkit to supports the development of the transformational strategy of organizations.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] L. Whelan, "Measuring the Success Factors of Strategic Design Implementation" IEEE Technology and Engineering Management Conference Europe 2021 <https://doi.org/10.1109/TEMSCON-EUR52034.2021.9488632>.
- [2] T. Brown, "Change by Design" Journal of Product Innovation Management 2019 <https://doi.org/10.1111/j.1540-5885.2011.00806.x>
- [3] R. Grant, Knowledge Management Theories. In: Augier M., Teece D.J. (eds) The Palgrave Encyclopedia of Strategic Management. Palgrave Macmillan, London.2018 https://doi.org/10.1057/978-1-137-00772-8_492
- [4] GIPC Global Innovation Policy Centre , Chamber of Commerce. Washington DC 2021 URL: <https://www.theglobalipcenter.com/>
- [4] D. Runde, "The OECD Faces a Decision Point in 2021" Centre for Strategic & International Studies. Washington DC 2020 URL: <https://www.csis.org/analysis/oecd-faces-decision-point-2021>
- [6] EGFSN. Winning by Design. Dublin: Expert Group on Future Skills Needs. National Skills Council, Dublin 2017 URL: <https://enterprise.gov.ie/en/Publications/Winning-by-Design.html>
- [7] J. Goddard, Connecting Universities to Regional Growth. Brussels: European Commission. 2011 doi:10.1093/cje/bes005 http://ec.europa.eu/regional_policy/sources/docgener/presenta/universities2011/universities2011_en.pdf
- [8] R. Verganti, Design Driven Innovation. The Oxford Handbook of Innovation Management 2014 DOI: 10.1093/oxfordhb/9780199694945.013.006
- [9] G. Calabretta, Strategic Design. Amsterdam: BIS Publishers 2018 ISBN 978 90 6369 445 6
- [10] E. Wenger, Cultivating Communities of Practice. Harvard Business Review 2002. ISBN 1-57851-330-8
- [11] G. Walsham, G., Interpreting Information Systems in Organization. Wiley New York. 1994 <https://doi.org/10.1177/017084069401500614>
- [12] M. Henry. The Rise and Fall of Strategic Planning. New York 1994 : Toronto :Free Press ; Maxwell Macmillan Canada
- [13] M. Pankina, "Axiology and Praxeology of Design Thinking. Knowledge E. Ural Federal University Russia 2020 DOI: 10.18502/kss.v4i11.7559
- [14] C. Frayling, "Research in Art and Design. London: Royal College of Art 1993. ISBN 1-874175-55-1