Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 3, 529-535 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Risk Management: The Case of Intrusion Detection using Data Mining Techniques

Ruba Obiedat[*]

King Abdullah the Second School of Information Technology, The University of Jordan, 11942, Jordan

ARTICLE INFO

Article history:
Received: 02 May, 2020
Accepted: 14 June, 2020
Online: 25 June, 2020

Keywords:
Risk Management
Intrusion Detection
Data Mining
Machine Learning

ABSTRACT

Every institution nowadays relies on their online system and framework to do businesses. Such procedures need more attention due to the massive amount of attacks that occurs. These procedures have to go first through the management team of the institution, in order to prevent exploits of the attackers. Thus, the risk management can easily control and identify the risk that occurs. One of these risks is an intrusion, which is an action or an act that the attacker invades someone's privacy to steal or damage their information. Various techniques have been proposed to prevent these actions in the literature. This research proposed an intrusion detection model to distinguish the most recent attacks using data mining techniques. Three machine learning classification models have been applied namely, J48, Random Forst and REPTree to improve the detection rate. Furthermore, a Feature Selection method has been applied in order to improve the effectiveness of the classifier and also overcome the high dimensionality which presents one of the main technical problems facing the intrusion detection systems and come up with the most important intrusion features affecting the system. These features can be very useful in protecting the systems from attackers. The results identify the top 11 effective features. The best results achieved by the J48 with a 76.271% accuracy rate.

1. Introduction

In the recent years, people all around the world become more depending on the information technology in all kinds, notably, with the expansion of the internet and automated businesses, and procedures from different fields [1]. People utilized computers and applications in order to acquire information about several things; such as stock costs, news, and online trade. On the other hand, others save the information of patient's medical records, credit card, and other personal data on their systems, either offline or online. Many organizations, for example, have a web presence as a fundamental structure of their businesses. Controlling and secure these critical assets and information that come from the decision of the management team [2, 3].

Consequently, without a good plan and scheme, risk can occur regularly. This kind of risk can harm the company for example, in a severe way, especially controlling the intrusion and detect each attack that happens [2]. The availability and integrity of the systems must be ensured against various threats, such as hacking or damaging, in which eventually can hurt the image of the institution. Hence, the secure information and the communication turned out to be indispensably vital. Furthermore, the need to detect privacy breaches and information security demands of a robust intrusion detection and prevention systems (IDPSs) are more necessity [2-5].

Several techniques have been proposed to prevent such vulnerability. The most recent and effective one is "Data Mining", which is a method of understanding and finding the pattern of the data [6, 7]. This data usually collected from different sources, each one of them portrays a case that occurs on different scenarios. In real life, data can portray as stones and sand, and mining these gravels to extract the jewelry (useful information). Therefore, extracting such benefit information can lead us to prevent attacks and leaking information in a more effective way. In our case, the dataset of these logs is collected, where the instances of each attack and non-attack portrayed as a row, while the columns are the characteristics of each instance. These characteristics called features. Another critical process can be used to help us knowing and understanding the pattern, and hidden information is called Feature Selection. It is a way to remove redundancy and not important features from the dataset and keep the most important ones without affecting the essential information of the data, it relies on identifying the features which are independent of each other but

[*] Ruba Obiedat, Email: r.obiedat@ju.edu.jo

https://dx.doi.org/10.25046/aj050365

highly relevant to the output at the same time using the goodness metrics. This technique can increase the accuracy of the classifier, utilize time and resources, reduce complexity, and support generalization [5, 8, 9].

This research presented a model using Data Mining techniques to identify these intrusion attacks. We used a public-online dataset that depicts these attacks. Then examined the data on different classifiers for evaluation and improving the detection rate of the Intrusion Detection System (IDS), which are; Decision Tree, RepTree, and Random Forest. And since IDS deals with massive amount of data [8], we then applied features selection to see the most important features that can help detect each attack. Feature selection method presents one of the most useful and commonly used techniques in data preprocessing step for the Intrusion Detection System (IDS). This technique is essential in our case here since the number of features in any Intrusion Detection System (IDS) is very huge whatever size of networks you work with and we need to reduce it. In addition, applying the feature selection technique can optimize the classifiers results and increase detection rate by removing unimportant features and help identifying the most relevant features affecting the system which present valuable information for the security teams as well.

The rest of this paper is organized as follows: section 2 presents the background of the intrusion detection history. The methodology produced in Section 3. Finally, section 4 and 5 introduces the experiments and results, and the conclusion, respectively.

2. Background

In order to define Intrusion Detection and Prevention Systems (IDPSs), it's important to know the events prior the detection procedures. The term attack is an action that exploits the vulnerability of a system to cause a loss of an information or access. Therefore, intrusion is considered a kind of attack, in which the intruder tries to obtain access or damage the system. In short, it is a process of observation, detecting and analyzing the performance that known as a violation of the information security policies. [10] Illustrates the intrusion detection as a way of detecting the cyber-attacks that take place on the computer networks using a certain framework known as Intrusion Detection System (IDS). This framework usually detects any abnormal patterns of the system.

There are two types of an attack, outsider attack, which is an attack that comes from the previous external origins. The other one is the Insider attack, an unauthorized/authorized inner user attempts to gain and abuse the non-authorized/ authorized entrance privileges to the system. Thus, we need to prevent such acts.

Intrusion Detection is one of these acts and known as a process of observing of the computer or networks for any unauthorized access, activity or modification [11, 12]. Intrusion Detection System helps to detect any intrusions on systems. Also, it helps to catch any unusual behaviors in different ways that covers and displays warnings when logging into a system. Another important procedure is that Intrusion Prevention System (IPS) detects system warnings in real time and prevents unauthorized access to the system by using specific software. Moreover, it can help to prevent any potential harmful events [11]. Intrusion Prevention System can detect threats in various ways; it can work with combination of access control (firewall/router) or any other security controls in the systems to prevent planned attacks. It can block the malicious network activity and configure or reconfigure privacy controls in browser settings to block these attacks [2].

However, shutdown the prevention characteristics in IPS cause them to operate as ID systems. Although IPS and IDS both monitor and analyze the network flow, IPS is recognized to be an expansion of IDS. IPS and IDS both are implemented to find unionize movement. The major duty of the IDS is to alert the unauthorized movements that are taking place. On the other hand, IPS is used for more effective protection by improving IDS and other common security solutions more. A powerful risk management process is essential part of security. Risk management has to determine the necessary security controls to decrease the danger to a suitable level by organizations. For effective IDPS design, it is necessary to obtain appropriate risk management-based requirements [2, 13].

2.1. Risk Management

Application, and network service of any computer and communication systems, usually behave normally and expected to be guarded against misuse, through a combination of privacy and safety and so on to be available and accessible. Similarly, these functions should provide a secure and trusted data transmission services to the end-users. Risk management means to control the risk and find a way to reduce the risk to an acceptable level. [14] determined the risk management as the procedure enables the mangers of IT to adjust the financial as well as operational expenses of protection and to satisfy the mission capacity by securing the Information Technology frameworks and information that eventually help their associations' missions [13, 14].

Security concerns can rapidly disintegrate client certainty and conceivably diminish the appropriation rate and rate of a degree of profitability for strategically critical items or administrations. Risk management procedure is a critical part of a successful security system. Main objective of an organization's risk management procedure must take into consideration the organization and its capacity to accomplish its central goal and missions, instead of essentially its IT resources and assets. Risk management shouldn't be treated as only a technical issue, but as a fundamental critical management issue inside organizations. Risk protection plans are described by understanding, identifying and accepting the leftover risks related to the use of information systems. In order to improve protection of the organizations from serious and increasingly threats of information systems, organizations should utilize a risk strategy alongside IDPSs, as a whole system of protection to secure the CIA triangle (Confidentiality, Integrity, Availability) of information systems [15, 16].

2.2. Intrusion Detection and Prevention Systems

Intrusion Detection System is used to observe any usual movements on a computer system or network. Moreover, it shows alerts on any possible violations, threats or attacks on computer system or network. Intrusion Detection System is both hardware and software application that work together to monitor and control computer system as well as computer network for any security policy abuse, furthermore, it reacts to any unusual activities by sending a message or alerts to the admin of the system in different ways [16]. Intrusion Prevention means to operate intrusion detection and try to stop any potential harmful events. Moreover, Intrusion Prevention System has the same capabilities of IDS of stopping potential accidents; it is designed to secure systems from any abuses or troubles. However, the main difference between IPS and IDS is that IDS warns in case of any attack attempt while IPS tries to block the attack from happening [17].

2.3. Important of the Intrusion Detection

Each organization utilizing computer systems must take into consideration the important of security. The fact that Information Security is an area which depends on specialists to enhance the safety of the enterprise most valuable resource can't be underestimated. The majority of businesses exclusively carry out high standard security policy arrangements, despite the fact that the largest threats are from inner sources. Furthermore, organizations perform network security arrangements which are intended to keep network assets safe. IDPSs improve the safety of InfoSec and network with denying the any unauthorized access of discovered attacks. Moreover, it can provide valuable clues on the best defense for the attack [8]. Unfortunately, there is nothing called "absolutely secure system", any system is vulnerable toward violation by insiders who misuse they system for their own benefits. In order to understand the requirements on focus and improve advanced IDPSs, an overlook into the numbers of incidents and data breaches must be studied [18, 19].

In the last decade, statistics record shows that the technological progress becomes more complex and security concerns grow more and more as shown in Figure 1 [20]. The incidents are limited for the danger and risk of privacy violations, security breaches, and malicious intrusions. Concurrently, with the increasing number of the computer systems as well as the huge expansion in the total number of threats, it has become even more aggressive than ever before, especially with the continuous increasing in size and speed of the networks. It is obvious that enterprises require intelligent InfoSec techniques such as IDPSs which can detect the recent formed attacks to limit the loss of data and keep up the privacy of data with a quick response.

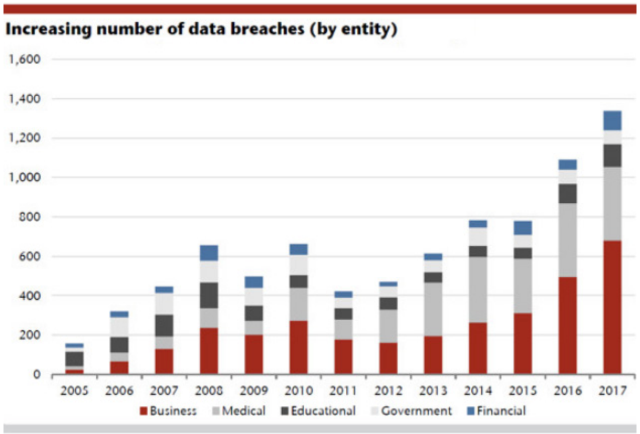

Figure 1: Numbers of data breaches

3. Related Work

Some researchers such as [21] introduced a framework which works with actual Intrusion Detection issues in classifying network data analysis into abnormal and typical behaviors. His paper suggests a "multi-level hybrid intrusion detection model" which uses learning machine as well as support vector machine to enhance efficiency of recognizing unknown and known attacks. Also, he added an updates K-means algorithm in order to develop a strong training dataset in order to improve the classifiers performance. The modified K-means is utilized to make new short training datasets to the whole original training dataset, essentially decrease the training time of classifiers, and enhance the performance of IDS. Analysis of the different techniques based on the similar dataset, the proposed model reveals and show high efficiency in accuracy and attack detection.

Furthermore, [6] proposed an "Internal Intrusion Detection and Protection System (IIDPS)", he suggested that in order to discover attacks, we need to utilize a forensic and data mining techniques. IIDPS generate users' individual profiles to remain track of users using their behavior as a forensic feature and decides if a user is the account owner or is not, by examining through user present PC usage practices with the patterns gathered in the account owner's personal profile. The results show that the user identification accuracy rate was 94.29%, while the reaction time is under 0.45 s, suggesting that it can manage to secure the system from insider attacks efficiently and effectively.

In addition, [22] presented a system to defeat menaces each time a detection arrangement was demanded because of highly spread in networks. Within the development of the arrangement, attackers have become stronger and each one damage network protection. Therefore, a need of the Intrusion Detection arrangement appeared to be a vital and necessary instrument in network security. Detection of such attacks loud intrusions normally rely upon the strength of Intrusion Detection Arrangement (IDS). In his approach, a number of elements have been coordinated for utilizing the methods; these systems have their very own advantages and disadvantages. In his paper, he focuses on various classification methods.

A detailed analysis and investigation of different machine learning methods have been developed, for locating the cause or problems related to different machine learning strategies in detecting the intrusive actions [23]. The results identified with low-frequency attacks utilizing network attack logs and dataset are as well discussed and viable techniques are recommended for development. Machine learning methods have been compared and analyzed in terms of their exposure ability for detecting the different classification of attacks. Limitations related to every classification of them are as well discussed. Different data mining devices for machine learning have as well been carried in his paper.

As can be seen in the mentioned works, many researchers focused on machine learning methods side to improve the detection rate of IDS but few had tried to investigate the impact of the features of the datasets. However, in this work, we study the impact of these features and what the implication of them, especially, the most important ones using feature selection method. Thus, improve the Detection performance.

4. Methodology

In this section, a description of the proposed approach processes is presented. As shown in Figure 2, five main processes are defined, namely, Data Mining, Classification Models, Data Description, Feature Selection methods and Evaluation.

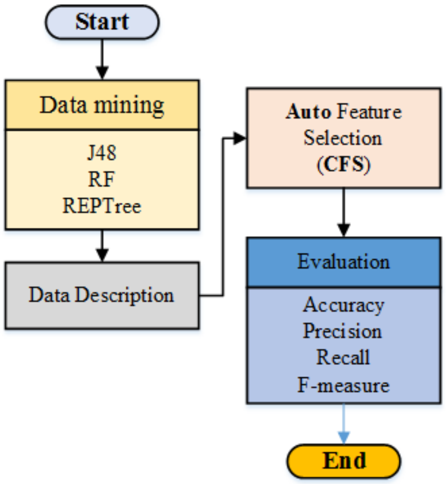

Figure 2. Methodology Steps

4.1. Data Mining

Data Mining reveals an important pattern, connection and directions through studying and analyzing each data using different machine learning techniques, data visualization methods, statistics, and artificial intelligence. This procedure effectively referred implicit, previously unknown, and possibly helpful information found in the data, therefore empowering the detection of patterns, the generation, and examination of hypotheses, and the creation of visualizations [24].

4.2. Decision Trees C4.5 (J48)

C4.5 is one of the algorithms related to decision tree which create trees from preparing data utilizing information gain and entropy. It is considered as an extension of the "ID3 algorithm" that overcomes most of its weakness. At every node, the algorithm checks the information gain and selects the attribute with the highest gain as splitting node, dividing the samples into subsets collection, the most significant characteristic value will settle on the choice then the procedure is taking place again upon smaller size of subsets [25].

4.3. Random Forest (RF)

This technique is based on the shell of a lot of decision trees, utilizing a stochastic procedure over the base of C4.5 algorithms. It was suggested by [26] and intends to make independent and uncorrelated trees dependent on the various and random input vectors attached by similar distribution. The outcome is the mean of the created trees through the process. RF is easy to implement, robust and able to handle large amount of data.

4.4. Reduced Error Pruning Tree (REPTree)

This algorithm is used to deal with decision tree noise. REPTree applies regression tree logic. It was presented in the post-pruning method founded on the concepts of [27, 28]. [29] defined the logarithmic procedure as: the initial step, where the data is separated into two groups. The first is the developing set, generated or created into a realization of learning algorithms, while the second is the pruning set that is generated through removing the subtree rooted at that node, until the outcomes in a predictive accuracy drop measured through the pruning set.

4.5. Data Description

In this work, the experiments applied on the intrusion-detection dataset that publicly available on the Kaggle repository/website [30]. The dataset consists of 20000 instances and 42 features. These features have different characteristics and methods in order to extract them, to name a few, duration, protocol-type, num-failed-logins, service, num-compromised, and so on. All features have numeric values except the protocol-type feature where it is a category-based.

4.6. Feature Selection methods

Feature selection (FS) is a method to choose the best group of features in order to use them in the developed model development. Overall, FS is designed by separating the redundant or irrelevant features without maintaining any loss of information while concentrating on the only relevant ones. It is one of the most vital goodness metrics to select features; generally, we say that a model is good if it is relevant to the output but not redundant with the other relevant features at the same time. The objective of the procedure is to reduce the time of training period of the classification models. Moreover, this procedure helps in developing the capacity of the model by decreasing the overfitting and to encourage researchers to facilitation their models and make them simpler to understand [31].

Feature selection techniques is an essential and effective step in data mining and has been adopted by many researchers in different data mining fields such as pattern recognition, network security and IDS. There are two types of feature selection approach; a Wrapper feature selection and Filter based feature selection. Wrapper method employs the performance of the machine learning algorithm to evaluate the usefulness of features, while Filtering approach relies on the characteristics of the training data itself to identify the redundant or irrelevant data [32].In this research we use one of the most common filter based feature selection method; which is Correlation-based Feature Selection.

4.7. Correlation-based Feature Selection

Correlation-based Feature Selection (CFS) is a heuristic type feature selection method that uses a search algorithm and evaluation criteria to select a subset of features. It is the most popular measure for the dependency between two variables. CFS measures each feature's goodness by predicting the usefulness of individual features to predict the class label and their intercorrelation level, and then come up with the optimal subset of features relevant to the class with no redundancy. In short, the explanation can be summarized: "Good feature subsets contain features highly correlated with the class, yet uncorrelated with each other" [32]. Therefore, the selection criteria in CFS select the best subset of features automatically. The following equation computes the merit of feature subset S with k number of features.

$$Merit_s = \frac{k\overline{r}_{cf}}{\sqrt{k+k(k-1)\overline{r}_{ff}}} \quad (1)$$

5. Evaluation

The Confusion matrix is a table that contains a summary of the prediction results for the classification system. Table 1 presents a confusion matrix for a binary classifier. It is used in order to assess the intrusion detection models of the current research and will be pointed out to as a primary source for the evaluation.

Table 1: Confusion Matrix

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Accuracy: determined by ratio of correctly classified instances of the two classes divided by the total number of instances.

$$Accuracy = \frac{TP+TN}{TR+TN+FP+FN} \quad (2)$$

Precision: the ratio of classes classified as attackers that actually are attackers

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Recall: the ratio of correctly classified intrusion attack divided by the total number of instances classifies as attackers.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

F-measure is determined as a weighted mean of the recall and precision.

$$F - measure = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (5)$$

6. Experiments and Results

In this section, we examined the data on several classification models. Named; Decision Tree, RepTree, and Random Forest Then we applied a feature selection technique to select the best subset of features for the classifiers and improve the accuracy of the results. In sum, first, we examine the data without applying the feature selection method. Then, we apply the feature selection before the examination on the classification model.

As shown in Table 2, the best accuracy obtained by J48 with 75.424%, whiles the second and third achieved by RepTree and RF, with 71.186% and 70.339%, respectively. For the recall measure, the highest results demonstrated by RepTree and RF at the same rate, 0.960%, and the J48 come third with 0.940. In precision and F-measure, J48 classifier has the best results with 0.671% and 0.783%, followed by RepTree with 0.623% and 0.756. The last classifier was RF with 0.600% for precision and 0.738% for the F-measure.

Table 2: Results of J48, RepTree and RF without Feature selection.

Data Without FS	**All features**			
	Acc	Recall	Precision	F-Measure
J48	75.424	0.940	0.671	0.783
RepTree	71.186	0.960	0.623	0.756
RF	70.339	0.960	0.600	0.738

The original data had 41 features and decreased to 11 features after applying the feature selection method as shown in Table 3. Table 3 contains the top11 features relevant to the attacks with a brief description about each feature taken from the Kaggle repository website. These data can be very useful in identifying important intrusion features and protecting the system from attackers. After finding the optimal subset of features we examine the classifiers again against these 11 features. As shown in Table 4, most of the results shows a brief increase, where the best classifier in the original data has a nearly 0.847 increase in accuracy with 76.271% rate — the second best results achieved by RF with 74.576%, unlike the original data where the second rank obtained by RepTree. The RF shows a notable increase with 4.237% which is more improved than the previous experiments, the least accurate of all classifier demonstrated by the RepTree with 72.034%. For the F-measure and precision, again the J48 had the highest results with 0.734% and 0.825%, the second and third classifier were RF and RepTree for both measures, respectively.

Table 3: Top selected features.

#	Features	Description
1	Service	Network service on the destination, e.g., http, telnet, etc.
2	Flag	Normal or error status of the connection
3	src_bytes	Number of data bytes from source to destination
4	dst_bytes	Number of data bytes from destination to source
5	Urgent	Number of urgent packets
6	logged_in	1 If successfully logged-in; 0 otherwise
7	srv_serror_rate	% Of connections that have "SYN" errors (same-service connections)
8	same_srv_rate	% Of connections to the same service (same-host connections)
9	diff_srv_rate	% Of connections to different services (same-host connections)
10	dst_host_srv_diff_host_rate	Diff_host_rate for destination host
11	dst_host_serror_rate	Serror_rate for destination host

Table 4: Results of J48, RepTree and RF with Feature selection.

Data WithFS	Selected features			
	Acc	Recall	Precision	F-Measure
J48	76.271	0.940	0.734	0.825
RepTree	72.034	0.940	0.644	0.764
RF	74.576	0.940	0.653	0.770

In sum, the feature selection method increases the results for all measures expect the recall measure. Decreasing the features from 41 to 11, not only shows a notable increase on the results but also, help us to know the best-selected features that can identify the intrusion detection attack form the non-attack class. As a result, we can conclude that the experimental results and the selected features show that the data mining techniques as classifiers and feature selection method can play an important role in improving risk management, helping the security teams to build their security strategies and prevent exploits of the attackers, which keeps their networks safe and secured. Finally, the risk factors can be further analyzed to study how they affect the intrusion detection as a case study of risk management.

7. Conclusion and Future Work

The current computer network is a risky domain, packed up with attackers that have millions of hours available to operate against the most grounded of security strategies. The best way to control them is to know when they are trying an attack and counter their efforts. Choosing the right intrusion detection system is the key to ensuring that an enterprise's networks and systems stay secure. While security cases turn out to be more numerous, network intrusions become a significant threat, as a result IDPS is becoming increasingly important and necessary. These intelligent IDPSs should utilize several intelligent methods from the matter fields of data mining, machine learning, and artificial intelligence to support them to figure out what qualifies as an intrusion. This paper proposed an intrusion detection model to prevent and decrease the intrusion attacks. Three machine learning classifiers were applied; which are J48, RepTree and RF to check the detection rate. Moreover, as the network traffic is becoming very huge with massive amount of data, leading to increased number of redundant and irrelevant features which decrease the IDS detection rate, consuming more resources, slowdown the detection process and increase complexity. Consequently, a feature selection method has been applied using the Correlation-based Feature Selection in order to find out the most relevant features to the detection process and remove unnecessary or redundant one. This approach is not only able to reduce the time consuming of the experiments but also increase the results and lead to higher accuracy. The best obtain accuracy achieved by the J48 with 76.271% using the top 11 features identified by the feature selection method. For future work, it is aimed to increase the number of the classifiers and uses different data to help us study the features in detail using a different technique of the features selection.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Atasoy, H., "The effects of broadband internet expansion on labor market outcomes", ILR Review, 66(2), 315-345, 2013.https://doi.org/10.2139/ssrn.1890709

[2] Michael E. Whitman, Herbert J. Mattord, Management of Information Security. Cengage Learning, 5th edition, 2016.

[3] Sennewald, C. A., & Baillie, C., Effective security management, Butterworth-Heinemann, 2020.

[4] Lewis, T. G., Critical infrastructure protection in homeland security: defending a networked nation, John Wiley & Sons, 2019.

[5] Amiri F, Yousefi M, Y, Lucas C, Shakery A, and Yazdani N., "Mutual information-based feature selection for intrusion detection systems", Journal of Network and Computer Applications, 34(4), 1184–1199, 2011.https://doi.org/10.1016/j.jnca.2011.01.002

[6] FY Leu, KL Tsai, YT Hsiao, CT Yang, "An internal intrusion detection and protection system by using data mining and forensic techniques", IEEE Systems Journal, 11(2), 427-438, 2015.https://doi.org/10.1109/jsyst.2015.2418434

[7] Leskovec, J., Rajaraman, A., & Ullman, J. D., Mining of massive data sets, Cambridge university press, 2020.

[8] Yang Li, Jun-Li Wang, Zhi-Hong Tian, Tian-Bo Lu, Chen Young, "Building lightweight intrusion detection system using wrapper-based feature selection mechanisms", Computers & Security, 28(6):466-475, 2009.https://doi.org/10.1016/j.cose.2009.01.001

[9] Ayman I. Madbouly Amr M. Gody ,Tamer M. Barakat, "Relevant feature selection model using data mining for intrusion detection system", International Journal of Engineering Trends and Technology (IJETT), 9 (10), 2014.https://doi.org/10.14445/22315381/ijett-v9p296

[10] D.E. Denning, "An intrusion-detection model", IEEE Transactions on software engineering, (2), 222-232, 1987.https://doi.org/10.1109/tse.1987.232894

[11] Lee, W., &Stolfo, S. J., "Data mining approaches for intrusion detection" in 1998 Proceedings of the 7th USENIX Security Symposium, San Antonio Texas, USA, 1998. https://doi.org/10.21236/ada401496

[12] Vidal, J. M., Monge, M. A. S., & Monterrubio, S. M. M., "Anomaly-Based Intrusion Detection: Adapting to Present and Forthcoming Communication Environments", In Handbook of Research on Machine and Deep Learning Applications for Cyber Security, pp. 195-218, IGI Global, 2020.https://doi.org/10.4018/978-1-5225-9611-0.ch010

[13] S. Kamiya, J. Kang, J. Kim, A. Milidonis & R. Stulz, "Risk management, firm reputation, and the impact of successful cyberattacks on target firms", Journal of Financial Economics, 2020.https://doi.org/10.1016/j.jfineco.2019.05.019

[14] Chichakli, R., "Information systems risk management", 2009.

[15] Malik, M. F., Zaman, M., &Buckby, S., "Enterprise risk management and firm performance: Role of the risk committee", Journal of Contemporary Accounting & Economics", 16(1), 100178, 2020.https://doi.org/10.1016/j.jcae.2019.100178

[16] Whitman, M. E., & Mattord, H. J., "Readings and cases in the management of information security: Law and Ethics", Information Security Professional, 2005.

[17] Martin, C., What is IPS and how intrusion prevention system work, available online on: https://www.paloaltonetworks.com/cyberpedia/what-is-an-intrusion-prevention-system-ips, 2009.

[18] Beal, V., Intrusion detection (IDS) and prevention (IPS) systems, 2005.

[19] Sundaram, A., "An introduction to intrusion detection", Crossroads, 2(4), 3-7, 1996.https://doi.org/10.1145/332159.332161

[20] Reklaitis, V., "How the number of data breaches is soaring - in one chart", Retrieved April 20, 2020, fromhttps://www.marketwatch.com/story/how-the-number-of-data-breaches-is-soaring-in-one-chart-2018-02-26, 2018.

[21] Al-Yaseen, W. L., Othman, Z. A., & Nazri, M. Z. A., "Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system", Expert Systems with Applications, 67, 296-303, 2017.https://doi.org/10.1016/j.eswa.2016.09.041

[22] Sree, S. B., Kernel Based Intrusion Detection Using Data Mining Techniques, 2018.

[23] Mishra, P., Varadharajan, V., Tupakula, U., & Pilli, E. S., "A detailed investigation and analysis of using machine learning techniques for intrusion detection", IEEE Communications Surveys & Tutorials, 21(1), 686 – 728, 2019. https://doi.org/10.1109/comst.2018.2847722

[24] Hand, D. J., "Principles of data mining", Drug safety, 30(7), 621-622, 2007.https://doi.org/10.2165/00002018-200730070-00010

[25] Quinlan, J. R., C4. 5: programs for machine learning, Morgan Kaufmann Publisher, 2014.

[26] Breiman, L., “Random forests”, Machine learning, 45(1), 5-32, 2001.https://doi.org/10.1023/a:1010933404324

[27] Pagallo, G., & Haussler, D., “Boolean feature discovery in empirical learning”, Machine learning, 5(1), 71-99. 1990, https://doi.org/10.1007/BF00115895.

[28] Quinlan, J. R., “Simplifying decision trees”, International journal of man-machine studies, 27(3), 221-234, 1987, https://doi.org/10.1016/S0020-7373(87)80053-6.

[29] Fürnkranz, J., &Widmer, G., “Incremental reduced error pruning” in 1994 Proceedings of the Eleventh International Conference, Rutgers University, New Brunswick, NJ, USA, 1994. https://doi.org/10.1016/b978-1-55860-335-6.50017-9

[30] Intrusion detection, Retrieved 29 April 2020, from https://www.kaggle.com/what0919/intrusion-detection, 2020.

[31] Hee-su Chae, Byung-oh Jo, Sang-Hyun Choi & Twae-kyung Park, “Feature selection for intrusion detection using NSL-KDD”, Recent advances in computer science, 184—187, 2013.https://doi.org/10.1109/icomitee.2019.8920961

[32] Chen,Y., Yang Li, Xue-Qi Cheng, & Li Guo., “Survey and taxonomy of feature selection algorithms in intrusion detection system”, In 2006 International Conference on Information Security and Cryptology, pages153–167, 2006.https://doi.org/10.1007/11937807_13