ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 6, No. 2, 378-386 (2021)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Ontology Based Privacy Preservation over Encrypted Data using Attribute-Based Encryption Technique

Rubin Thottupurathu Jose[1,*], Sojan Lal Poulose[2]

[1]*Department of Computer Sciences, Amal Jyothi College of Engineering, Kanjirapally 686518, India*

[2]*Department of Computer Sciences, Mar-Baselious Institute of Technology and Science, Kothamangalam 686693, India*

ARTICLE INFO

Article history:
Received: 06 January, 2021
Accepted: 08 March, 2021
Online: 17 March, 2021

Keywords:
Attribute Based Encryption
Cloud Computing
Data Owner
Ontology
Privacy

ABSTRACT

The web documents are automatically interacting to discover the information by web mining, which is one of the applications of Cloud Computing (CC) technologies. These documents may be in the form of structured, semi-structured, or unstructured formats. In current web technologies, the Semantic Web is an extension for better enabling the people and computers to work together, where the information is well defined. Before storing the data to the cloud server, data owners should encrypt their data for privacy and security concerns. At the same time, the end-user, who is finding the data related to specific keywords, suggests the research on searchable encryption technique. In this research work, fine-grained authorization of search was achieved by developing the Attribute-Based Encryption (ABE) search technique, which is under the distribution of multiple attribute authorization. Finally, to validate this approach, an experimental study is conducted on Wikipedia as an ontology with existing techniques. This research applies the Attribute based encryption and search method for the effective search and improve security in the cloud. Access policy, cipher text and secret key is developed based on the Attribute selected from the data. The Lagrange interpolation method is applied for the search process and registration key is applied to access the data. The privacy preserving efficiency of the proposed model is 99.2 % and existing Hierarchical-ABE method has 96 % efficiency.

1. Introduction

In the past few years, the World Wide Web (WWW) is considered as one of the most important resources for knowledge discoveries and retrieval of useful information due to the vast amount of available online data [1, 2]. The process of retrieving the information from knowledge discovery by using web mining technologies is one of the right solutions on Web. The performance is improved for Web based data warehousing, web information retrievals and question answering by extracting the knowledge from the Web [3]. There is an availability of user's private sensitive data in web based databases because the web services are integrated into our daily life. These sensitive data of user can be accessed by unauthorized persons who are having malicious intentions and the reason behind this is the huge availability of web [4]. The requirements of the users should be matched with the privacy policies of service provider to reduce the disclosure of private data. According to the various policies such as privacy as well as security, the service users' privacy sensitive data will be secured [5, 6]. CC offers benefits and more services to system software and hardware of data center over Internet [7].

According to the customer requirements, the inclusive platforms or apparatus are divided, then the software is delivered to end user. In addition to this, the communication and storage over internet are considered as some of the important services of CC. Also, three important protection issues namely accessibility, privacy and reliability can be tackled by the techniques of CC [8]. A rich amount of information can be generated by web service users, while the service providers' websites are browsed and accessed through social networking sites for posting the comments and reviews of products by end user, then their data are stored in the cloud [9]. The new threats for digital life of user's privacy are raised due to increases in online activities of user and technologies [10]. The service users signed to the privacy policy of providers unknowingly for providing the authorization to them for collecting and sharing the personal identification information of user while accessing the web services [11]. When accepting the policies of providers, some users thought that they secured their privacy

*Corresponding Author: Rubin Thottupurathu Jose, Email: rubinthottupuram@gmail.com

information, but their rights are surrendered to the providers [12, 13]. This research work aims to reduce the disclosure of privacy related information by using prevalent semantic web based technologies. There are two major contributions of this work such as encrypt the collected data using ABE technique to preserve the users' access rights to the data. Then, the Attribute-based Search (ABS) scheme is developed on encrypted cloud data to support multiple users and multiple data owners. In order to reduce the search space of documents, the second index are exploited by cloud server during the process of search and the user can able to access these documents. Then, the relevant documents are retrieved by using the first index and these two methods are used as an effective and secure access control policy, where the data owner can outsource the encrypted data and retrieves the secure documents over the cloud. The proposed ABE search technique properties are as follows:

- This research involves in applying the fine-grained authorization with Attribute based Encryption and search algorithm method for effective search and improve security. The Attribute is used to develop cipher text and secret key for the user. Access policy are developed based on cipher text to control access over the data.
- The Lagrange interpolation method is used for the search process and registration key is used to access the data and Two index is used in this method for effective search and reduce the computational time for the authorization.

The proposed ABE search technique is compared with existing models in encryption and search process to analysis the performance. The performance evaluation is carried out for various parameters such as precision, recall, auditing time etc. The result shows that the proposed model has higher efficiency of 99.2 % and existing HABE method has 96 % efficiency shows lower auditing time compared to existing method.

The rest of the paper consists of a survey of recent techniques which is used to retrieve the encrypted data in semantic web mining are described in Section 2. The proposed techniques for preserving the users' access rights are presented in Section 3. The validation of the proposed method with existing techniques in terms of various metrics are represented in Section 4. Finally, the conclusion of this research work with further development is discussed in Section 5.

2. Literature Review

In this section, a review of recent techniques namely encryption techniques, access control solution and disclosure discovery methods based on privacy-preservation of end user tare presented in [14-20].

In [14], the author implemented a Hierarchical-ABE (HABE) based modular padding scheme to address the challenges in public auditing. The semantic ontology was generated by assigning the key parameters to various data levels. The public auditing was carried out by this generated ontology, then the method verified the types of the user request and modular padding was performed. The efficiency of data sharing and quality of public auditing was improved by the HABE method. In case of system failure, the retrieval of user data was not concentrated by this HABE method.

In [15], the author developed a Three-Fold Sanitization (TFS) framework for detecting the sensitive topic semantically. The Gibbs Sampling was used to recognize the sensitive topic clusters with high location entropy, which was subjected to Semantic Sensitive Access Rule-LDA Topic Model (SSAR-LDA). The sensitive terms were eliminated by enhancing the privacy preserving policy of TFS, which replaced that terms with appropriate hierarchical generalization terms. The method faced the problem of language ambiguity, which leads to poor performance in detecting user profiles.

In [16], the author implemented the ontology-based access control solution to overcome the problem of interoperable exchange of security policies and stealing authentication credentials. According to the estimated trust degree of user's request and criticality of data resources, the method encompassed a risk-aware authentication scheme to overcome the vulnerabilities of malicious activities. The method used the pseudonyms by developing the privacy-preserving authentication and reputation management which was used to avoid the exposing personal information of users. The method failed to focus on possible issues such as data sovereignty for providing privacy and security guarantees in cloud computing.

In [17], the author addressed the inefficiency of technical knowledge and rigidness of access control mechanism, a transparent and dynamic Privacy-Driven Access Control (PDAC) was developed for textual messages. According to the privacy requirements of publishers, the sensitive information was detected by assessing the semantics of messages automatically. There is no need for administrative efforts at the publication time because privacy enforcement was transparent both to publishers and readers. The semantic coherence of the protected messages was affected by improperly disambiguated terms, which was considered as the main limitation of this approach.

In [18], the author prevented the illegal disclosure of user's sensitive privacy information by developing a Private Data Chain Disclosure Discovery (PCDD) method. The cost and similarity degree of disclosure of private data were measured, then key private data and disclosure chain were detected based on the measurement of private data (i.e. cost and similarity degree). When user released the private data to other software services, the disclosure of user's sensitive private data was identified in time by using this approach. In the description tree, when there was a presence of huge quantity of un-matching nodes, then the time efficiency of this algorithm was very high.

In [19], the author proposed a Secure Inverted Index (SII) using homomorphic encryption and dummy documents technique to solve the privacy issues of end user. Even though, the two techniques had limitations, this approach obtained benefits from these two methods by using compressed table of encrypted scores and double score formula. The user's access rights to data were managed by using second secure inverted index. After encrypting the index documents, the eight dummy documents were added to test the search time of secure inverted index technique.

In [20], the author implemented an Index Hash Table (IHT) with Paillier Encryption (PE) for dynamic public auditing by recording the data property that was located at TPA. The information were migrated to TPA from CSP for the reduction of

computation cost and communication overhead. The privacy preservation was supported by combining the random masking with homomorphic authenticator, which was based on public key. When compared with previous scheme, the DHT method reduced the storage costs, communication and computation cost by achieving secured auditing in clouds.

In [21], the author developed a key-insulated ciphertext policy ABE with key exposure accountability (KI-CPABE-KEA) for providing protection of KE and achieving the user accountability. The ciphertext was decrypted by data receiver, when the self-centric policy was matched with the attributes of data receiver. Every private key of user had unique identifier, so system manager pinpointed the user's identity, if the private key was exposed for illegal data sharing. The security analysis showed that the KI-CPABE-KEA had higher efficiency for sharing the data in CC. However, the developed method was insufficient for data authentication in attribute based cryptosystem.

In [22], the author solved the issues of revealing the encryption's privacy by implementing the CP-ABE scheme. The encryptor's and decryptor's privacy policy was preserved by adding the hidden access polity with CP-ABE scheme. Before decryption process, the unnecessary operations were avoided and solved the efficiency issues by introducing the testing phase in the CP-ABE scheme. While comparing with decryption computation, the testing phase's computation cost was minimized and this developed scheme was secure under chosen-plaintext attack. According to AND gates with wildcards on multi-valued attributes, the access policy of the scheme was highly expressive and it was affected by decisional bilinear Diffie-Hellman (DBDH) attack.

In [23], the author protected the data confidentiality, while sharing the data over cloud by developing the CP-ABE with fine-grained access control. Under the standard mode, the assumption of DBDH was minimized by the security of the developed scheme. The security analysis showed that the developed scheme has higher efficiency and protected the user's privacy. The validation was conducted by using the parameters such as storage cost and computation cost. But, the developed CP-ABE scheme was only used for securing the data.

The existing techniques are implemented to secure only the encrypted data and failed to effectively retrieve the document. An objective of the research is to address the issues of sensitive information from the attributes such as HIV, Christianity, Berlin, Exact, Plugin, Subsume, Homosexuality, leakage of user data and to improve the efficiency in privacy preservation. According to the nature of the sensitive topics, the requirements are considered for the protection which will focus on protecting the sensitive data such as HIV (sensitive disease) Homosexuality (sexuality) and in order to protect against the attribute disclosure in identifying data (locations, name of the person) will be protected against identity disclosures. Similarly, Christianity includes sensitive information (religion, Belief etc.,), Berlin includes the Census data related with (location name or the city name) that are also protected against the identity disclosure.

The research work implements the ABE-based search algorithm for retrieving the related documents with higher accuracy.

3. Proposed Methodology

The sensitive information should be preserved for protecting the privacy of the end user. There may be a presence of contradictions in semantic web mining like extracting the useful information from the collected data without compromising the privacy of clients. While maintaining the usability of data, the privacy of sensitive data are preserved by developing various techniques namely differential policy, data perturbation and anonymization. But, the sensitive information are not mined from the given data context. Therefore, this information should be marked manually by data administrator, which is time consuming and tedious process. To address the issue, this research work implements the attribute based encryption and search algorithm (i.e. ABE and ABS) to preserve the privacy information of end users. In this section, a system model, threat model with design goals of proposed method are explained as below.

3.1. System Model

The system model for proposed ABE-based search technique is shown in Figure 1, which consists of four entities namely cloud server, data owner, administration server or Third Party Auditor (TPA) and data user. A set of files are encrypted and then uploaded to the cloud server by data owners. According to files and keywords, a secure searchable index is constructed by data owner using efficient ABE-based search technique and then these searchable indexes are submitted to the TPA by owner.

Suppose, a new user wants to search, initially he/she should register on the TPA server for searching the keyword of the data that is stored in the cloud. The user can able to send the authenticate attribute request to trusted authorities called Certificate Authorities (CA) by using legal identity, once the registration is successfully finished. The major function of CA is to verify the request and if the process of authentication is successful, then CA can provide the attribute key to end user. The encrypted trapdoor with searchable keyword can be generated by end user, when the keys are provided to them. The TPA re-encrypts the user's trapdoor with searchable keyword by using his secret keys. Then, TPA should submit the obtained index and trapdoor to the cloud server. While receiving the index and trapdoor, the server verifies the access control structure that belongs to the attributes of user's trapdoor. If the verification is successful, the cloud server returns the relevant cipher text by searching the encrypted index. The threat models are discussed below.

3.2. Threat Models

There is various important search scheme designed over the encrypted data using the threat model are discussed in this section.

Content Protection: Before outsourcing, the contents includes queries, indexes, and documents to cloud. The user should encrypt all these data.

Secure the Keyword: The frequency distribution and inter-distribution of a document are used by the proposed method to secure the data by hiding the information from the server. For every given document, the frequency distribution of a given term is described by term distribution and the distribution of term score

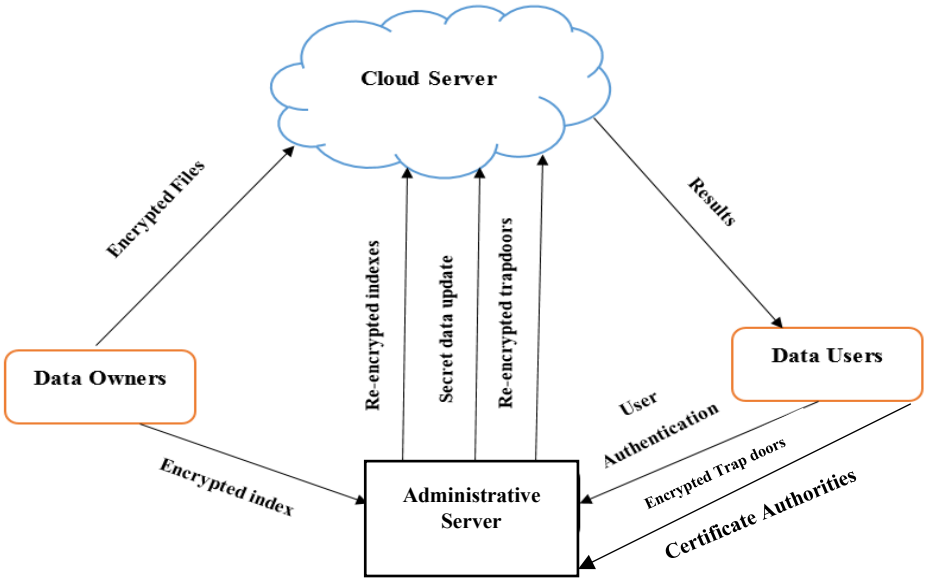

Figure 1: System Model of Proposed Method

is represented by inter-distribution. The links between documents and terms are prevented from the server by hiding these two features.

Trapdoor Unlinkability: The relationship between the collection of encrypted queries are gathered to prevent the information from the server, which can be carried out by proposed scheme. Therefore, there should be a non-deterministic of cryptosystem should be used.

Pattern Accessing: During the search, the proposed scheme should hide the sequence of resultant data from the server, while sending the data to user.

3.3. Proposed Methodology of Attribute Based Encryption Technique

Wikipedia is defined as ontology in this research work for exploring the meaning of documents and queries during the search process. There are more than four million English pages present in this ontology representation, where this research work chooses the Wikipedia due to its rich representation of information. Moreover, the articles in most languages and areas are presents in this ontology and the set of concepts are mapped between the concepts and terms, which are represented by query and document. In Attribute-based encryption, the ciphertext and secret key of a user depends on the attributes, where decryption process is carried out only when the attributes of ciphertext is matched with the user key. In this research study, the access policy are applied by data owner to control the access over the data collection by using ciphertext-policy ABE technique. During the encryption process, an access structure are defined for constructing the every data which will be included in the cipher text and the user's private key are presents in a set of attributes. If the access policy of the data are satisfied by the private key attributes, then the cipher text can be decrypt at the time of decryption process. Suppose, if it is not satisfied, then the user's private key will not able to decrypt the cipher text and also does not have the rights to access the data. The Table 1 shows the notation of parameters that are used in the research study.

Table 1: Parameter Notations

Notations	Definitions
MsK	Master Key
PuK	Public Key
At	Set of Attributes
PrK	Private Key
M	Number of Messages
As	Access Structure
Prk'	Set of New Private Key
A't	Set of New attribute set
CT	Cipher Text
K	Number of documents
Rki	Registration Key

There are five algorithms presents in this encryption method, which includes setup the process, generating the key, encrypt the data, decrypt and then delegate the data over the collected data. In the below section, these five algorithmic steps are discussed as:

- $Setup() \rightarrow \{Puk, MsK\}$: A two keys such as master key, which are represented as MsK and public key as Puk are generated by using this algorithm.
- $KeyGen(MsK, At) \rightarrow \Pr K$: This algorithm gives the input for encryption process by using two inputs, which includes a set of attributes At and MsK . Moreover, the private key for users as $\Pr K$ are generated by using this algorithm, where these keys are generated only for authorized users who are related to the attributes set At .
- $Enc(PuK, \mathrm{m}, \mathrm{AS}) \rightarrow CT$. A message m is encrypted by using a public key PuK under an Access Structure As in this algorithm, which are executed by the data owner.
- $Delegate(\Pr K, A't) \rightarrow \Pr K'$. This algorithm uses the two kinds of inputs, which includes attributes set as $A't$ and a private key as $\Pr K$. A $A't$ is included in At because a private key of user is related with a set of attributes At . A new

private key as $\Pr K'$ are developed by this algorithm that is the main aim of delegate, where this new private keys are associated with the set of attributes $A't$.

- $Dec(\Pr K, CT, PuK) \to m$. A cipher text CT are decrypted by user, where this algorithm is used to execute this process. An access structure As are presents in the cipher text CT, which are taken as an input by this algorithm. In addition, the public key PuK and a set of attributes At are also presents in the user's private key $\Pr K$. In order to get back the original message m, the cipher text CT should be decrypted, when the access structure As are satisfied by the set of attributes At.

The user's privacy information are preserved from the unauthorized access by using this encrypted data from the ABE technique. Then, cloud server will obtain the secure indexes and collection of encrypted data. A set of public keys are constructed by data owner, where trapdoors are built by these keys and finally the documents are decrypted by using private keys. The users who are authorized to the data, will receive these keys. A query is formulated by user to perform the search over the encrypted data using a public keys, which is used to build the trapdoor and these are send to the cloud server. The relevant documents are retrieved by using the search index, while the cloud receives the trapdoor function and then, authorized users obtained the $top-k$ documents. At last, the trusted authority will decrypt, filter and sorted the returned results at the user side. The next section will explain the ABS technique which is used to retrieve the documents for end user.

3.4. Proposed Method of Attribute Based Search Technique

An ABE technique encrypted the keywords and files of various data owners, and then the encrypted data are stored in the cloud server. Moreover, the secret data can be stored by administration server as TPA on this cloud server. Once the query request of user are received, all these data owners' files are searched by cloud server and then recalculate the Lagrange interpolation [24] (i.e. $L'R = e(RK_i, g_2)^{\beta}$, where RK_i is the registration key. The cloud processes the search request in two steps such as initialization process and retrieval process.

Initialization Process: i) Compute whether the encrypted index of first index with secret key should be equal as the encrypted index of second index with other secret key and justify the values of Lagrange interpolation. ii) When the cloud server obtains the trapdoor and searchable index, the user should compute encrypted trapdoor with his secret key. If the condition satisfies, the top-k documents can be retrieved from the cloud and delivers to the user.

Retrieval Process: The encrypted documents are searched by using specific keywords, which are explained in this phase are as follows:

i. Initially, the trapdoor function are called by creating the encrypted query, where this process can be done by an authorized user. Upon receiving the trapdoor, the cloud utilizes the first index to retrieve the relevant documents, and simultaneously it exploits the second index to get the list of documents IDs that the user has the right to access.

ii. The document IDs with their encrypted indexes are returned to user by launching the search function, when the encrypted queries are received by server.

iii. At last, the Sort function is used to filter and sort the returned documents in the user side. Then, the selected concepts are sorted with regard to their first index then based on their second index in the case of equality. Finally, the top concepts with their associated index are used to represent the document as a sort function

During the search process, the cloud server exploits to reduce the search space to documents accessible by the user and uses the key words to retrieve the relevant documents. Once the data retrieved from the database based on the query given by the end-user, the user need to verify whether it is relevant to the query or not. The next section will be described the validation of proposed method with other existing techniques.

Algorithm for the proposed ABE search technique

G_1, G_2 are bilinear group of order (p - prime), (g – generator group) G_1

$G_1 \times G_2 \to G_2$ is bilinear mapping d being the threshold value

Generate the public key and master key by randomly selecting the trusted center $t_1, \dots t_n$

y from finite field Z_q calculates the public key $P_k = (T_1 = g^{t_1}, \dots, T_n = g^{t_n})$

Generate Master Key $M_k = (t_1, \dots t_n, y)$

Generate private keys $D = \{D_i = g^{(q(i))/t_i)}\} \forall i \in AU$

Encrypt the message encrypted using set of attributes A_{CT} from $M \in G_2$ using a set of attributes A_{CT} and a random number $s \in Z_q$

CT = $(A_{CT}, E = MY^s = e(g,g)^{ys}, \{E_i = g^{t_i s}\} \forall i \in AU$

The encrypted data are supplied to the input of the decryption algorithm, and the output of the algorithm is obtained decrypted message.

If $|A_U \cap A_{CT}| \geq d$

Select d attributes

Compute the first and second index

Compute sort function

Searchable encryption is completed with security.

4. Result and Discussion

In this section, the implementation of the proposed method using tools are briefly described and the experimental validation of these approach with other existing techniques along with the results are evaluated and discussed.

4.1. Parameter Evaluation

The experiments are implemented using Python 3.7.3 on a computer with Intel Core i5 CPU 2.2 GHz with 8.00 GB RAM. In

this experimental analysis, the performance of ABE-based search technique is validated by using various parameters such as precision, recall, searching time and privacy preserving efficiency. In this research study, two objectives are considered, where the performance of two objectives needs to validate with various metrics. Here, the retrieval performance is validated by using precision and recall. The proposed ABE-based search technique is compared with existing techniques namely HABE [14], TFS [15], PDAC [17], PCDD [18] and SII [19], which are explained in below section.

Precision: Among the detected sensitive data nodes, precision predicts the correct number of sensitive data nodes. The mathematical expression for precision can be given in Eq. (1):

$$\Pr ecision = \frac{TP}{TP + FP} \tag{1}$$

where, TP represented as True Positive, FP represented as False Positive.

Recall: While calculating the total number of detected sensitive data nodes, recall gives the percentage of correctly identified sensitive data nodes, which is explained in Eq. (2):

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{2}$$

where, FN is described as False Negative.

4.2. Security Analysis

To analyze the security aspect of the proposed ABE scheme, the security analysis are examined whether the security controls that are presented in subsection 3.2 are respected. In addition, the security of the encrypted reverse index and the access rights of users are examined.

i) Protected content: This control encrypts data such as indexes, collected data and queries that passes through the cloud server. Initially, ABE technique is used to encrypt the data scores used in this approach, where identifiers is described as documents and concepts. A set of concepts with associated weights are illustrated as trapdoor. Then, ABE technique is used to encrypt the every weight and every concept, which are described by an identifier. The contents of documents are encrypted by the proposed method and enables to apply an access control policy. In conclusion, the developed approach respects the control of protected content because the indexes, databases and queries are encrypted.

ii) Privacy keyword: This restriction is to prevent the server from establishing a link between the terms and the documents. Two properties such as term distribution and inter-distribution are used for this purpose that must be hidden. On the one hand, the distribution of the term scores in a given document are presented by inter-distribution. This property is hidden by ABE technique, which enables encryption of scores in the reverse index. On the other hand, each document in the collection contains the frequency of a given word, which is presented by term distribution. This property is hidden by the dummy document technique, which prevents the server from knowing if a word belongs to the documents leading to the corresponding entry. From this, it is concluded that keyword privacy control is respected in developed approach.

iii) Trapdoor Unlinkability: This restriction is to prevent the server from linking between various trapdoors. For this purpose, the concepts (x) are selected at random from the set of ideas (y) that represent the query, for example, 10 x among the 100 y ideas is selected that represent the query. An ID is used to illustrate the each idea and ABE is used to encrypt its weights. This construction allows researchers to obtain various trapdoors for the same query. Therefore, the developed approach provides an undetermined encryption scheme, which allows to control the trap unlink ability.

iv) Access pattern: This restriction involves hiding user results from the server. While searching, a set of dummy document identifiers will always be provided with the correct result. The correct results are hidden from the server by using this false positives. In addition, developed technology allows the user to access the required documents without revealing their identities, which prevents the server from identifying false positives in the search result. Therefore, access pattern control is respected in this proposed approach.

4.3. Performance Evaluation of Proposed ABE Method

In this section, the validation of proposed ABE-based search technique is compared with PDAC [17] and PCDD [18] in terms of precision and recall. From the Wikipedia Ontology, this method chooses some sample sentences namely HIV, Christianity, Berlin, Exact, Plugin, Subsume, Homosexuality are used for validation. The ABE-based search method finds the access levels for sample sentences such as Infection and condition for HIV, religion and belief for Christianity, Location and City for Berlin, process and sexual activity for Homosexuality and so on. The security metrics involved in the proposed model are the privacy preservation efficiency, Verification time, auditing time and searching time. Table 2 shows the effectiveness of ABE-based search technique with existing techniques over precision and recall for sample sentences.

Table 2: Comparative Analysis of Proposed ABE-based search technique

Ontology Samples	**Methodology**					
	PDAC [17]		**PCDD [18]**		**Proposed ABE-based search technique**	
	Precision(%)	**Recall(%)**	**Precision(%)**	**Recall(%)**	**Precision(%)**	**Recall(%)**
HIV	66	77.77	65	63	80	87
Christianity	60	90	69	70	76	95
Berlin	75	80	73	59	84	86
Homosexuality	66	80	60	63	79	88
Exact	50	75	78	59	87	81
Plugin	57	71	73	63	91	79
Subsume	55	80	68	75	73.7	85.9

The above table explains the performance of proposed ABE-based search technique for precision and recall. The validated results stated that the proposed ABE shows better performance than PDAC and PCDD. The graphical representation for precision and recall is shown in Figure 2.

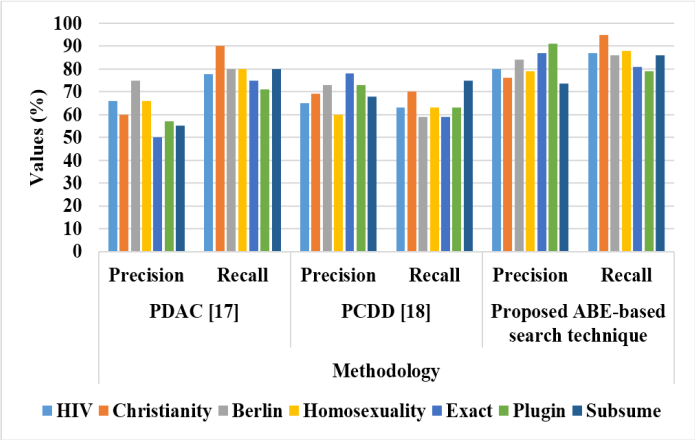

Figure 2: Performance of proposed ABE-based search technique

The existing technique PCDD achieved very low precision and recall for some samples such as HIV, Homosexuality, Berlin due to presence of noises and outliers in the Wikipedia ontology. These methods did not consider the user's access rights which leads poor performance in both precision and recall. The existing technique PDAC achieved higher precision and recall values, when compared with PCDD for all sample data. However, the disambiguated terms affect the semantic coherence of the protected message, which leads to low precision values in three samples namely exact, plugin and subsume (i.e. 50%, 57% and 55% precision). In this proposed method, the user access rights are preserved and also increases the semantic coherence of these messages by using ABS technique. Therefore, the validated results of ABE achieved nearly 80% in both precision and recall for all ontology samples. Table 3 shows the searching time of proposed method with existing techniques for retrieving the related documents using keywords.

Table 3: Searching Time of Proposed ABE-search based Technique

Methodology	Number of Queries							
	2	4	6	10	15	20	25	30
HABE [14]	24	29	34	40	51	57	64	70
TFS [15]	25	34	45	55	68	59	61	67
PCDD [18]	23	29	32	49	55	60	64	69
SII [19]	20	27	29	38	42	47	52	60
Proposed ABE-search based technique	18	23	26	31	39	44	49	54

The above table provides the searching time results in seconds for various number of queries. The searching time of proposed ABE is compared with various existing techniques namely HABE [14], TFS [15], PCDD [18] and SII [19] and their validated results are presented in the graphical structure, which is shown in Figure 3.

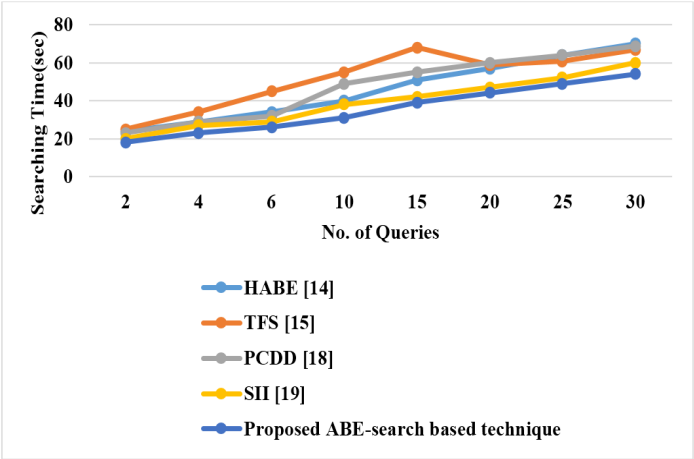

Figure 3: Searching Time of Proposed Method

Figure 3 shows that the proposed ABE-based search technique searches the queries in less number of time. When the sequential queries sizes increases, the searching time also increases. The existing techniques achieved nearly 70sec for 30th queries, but the proposed ABE method searches the sequential 30th queries in 54sec. Due to insufficient storage space, the existing techniques took larger time for searching the queries. The proposed ABE method overcomes the above issues by using index values. From the Table 2 and Figure 3 shows that the proposed ABE-based search technique performs better than existing techniques namely HABE, TFS, PCDD and SII. Table 4 shows the efficiency of privacy preserving of proposed method with various existing techniques, which was proposed in [14].

Table 4: Efficiency in Privacy Preservation of Proposed Technique

Methods	Efficiency (%)
Diffie Hellman	75
One-step	86
Multi-user inference	92
HABE	96
Proposed ABE-based search Technique	99.2

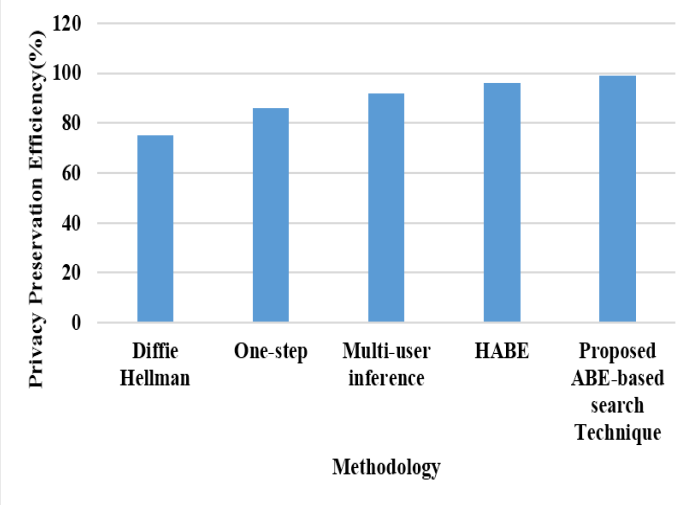

Figure 4: Efficiency of Privacy Preservation over Proposed Method

The experimental analysis of proposed method and their validated results are compared with various techniques, which is shown in Figure 4. The efficiency is used to calculate the privacy preservation of proposed ABE technique, which is stated in Table 4.

The existing techniques such as Diffie Hellman and one-step achieved very less efficiency (i.e. 75% and 86% privacy preservation in efficiency) when compared with other techniques. But, the HABE and Multi-user inference techniques achieved 96% and 92% efficiency due to padding process in the final step. The HABE method predicted the padding details from the ontology to generate the strings, then it adds more number of binary data to the end, which lead to time complexity and data loss from the server end. But, the proposed ABE-based search technique generates the strings when the encryption process takes place and stores the data in cloud. The proposed method avoids the data loss and preserved the privacy information of end user, which leads to achieve 99.2% efficiency. From the above experimental analysis of various parameters, the results stated that proposed ABE-based search technique achieved better results, when compared with existing techniques.

The experiments also evaluate the performance of ABE-based search technique in the verification scenario and compare it with DAP and Index Hash Table (IHT) with PE [20]. The verification time of the different block size is measured for the existing and proposed method, as shown in the Table 5 and Figure 5.

Table 5: Verification Time of ABE technique

Block Size (kB)	IHT - PA	DAP	ABE-based search technique
10	2.24	1.68	1.52
20	2.25	1.69	1.42
30	2.27	1.71	1.61
50	2.28	1.71	1.49
70	2.28	1.71	1.31
100	2.32	1.71	1.44
150	2.34	1.72	1.24
200	2.37	1.73	1.25

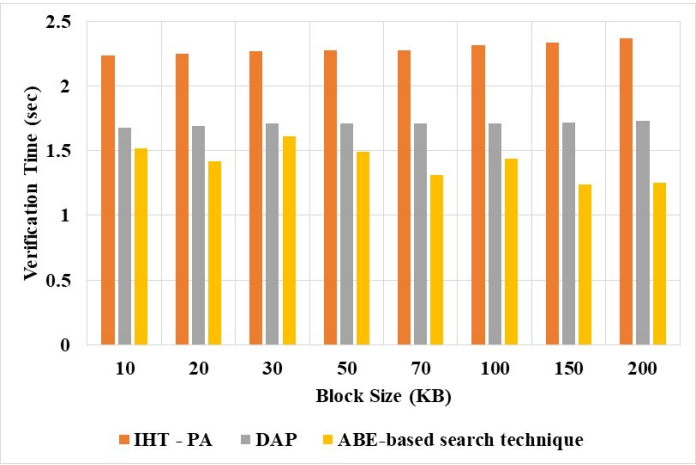

Figure 5: Verification of ABE-based search technique for different block size

The proposed ABE-based search technique has the lower verification time compared to the other existing method. The verification time of the DAP and ABE-based search technique methods has the much lower computational time than IHT-PA due to the significantly outweigh the disadvantage include by searching operation. When compared with other existing techniques, the proposed method achieved 1.25 sec for Block size 200. The experiments also evaluate the performance of ABE-based search technique in the batch auditing scenario and compare it with DAP and IHT-PE. The experimental results are as shown in Table 6 and Figure. 6.

Table 6: Auditing Time of Proposed Method

Block size(kB)	IHT - PA	DAP	ABE-based search technique
10	1.95	1.81	1.06
20	1.92	1.79	0.99
30	1.86	1.76	1.15
40	1.83	1.75	1.17
50	1.79	1.7	1.29
60	1.76	1.69	1.24
70	1.72	1.66	1.22
80	1.73	1.64	1.27
90	1.75	1.67	1.36
100	1.76	1.69	1.34

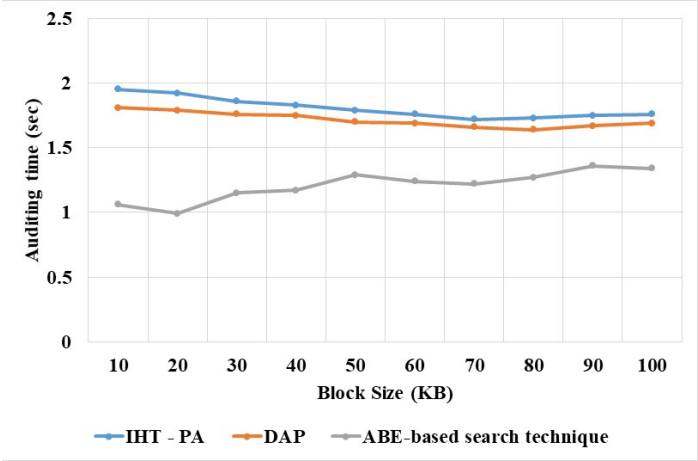

Figure 6: Auditing Time of Proposed Method

From the experimental results, it is clearly understood that the auditing time of proposed method is highly minimized with the existing techniques: IHT-PA and DAP. For instance, the proposed method nearly reduced the auditing time from 35%-40% for all files. In addition, the experimental results suggested that the batch auditing handle the verifications from multiple-users simultaneously. While performing the individual auditing for multiple times, this ABE-based search technique reduces the computational costs of TPA. Also, the results proved that the batch auditing protocol in ABE-based search technique is more efficient than that in DAP, and IHT-PA. The ABE-based search technique has lower computation time compared to the other existing method. The security of the method is increased by using

the secure key generation method. Hence, the proposed method can be applicable to practical use in the cloud auditing system.

5. Conclusion

The sensitive information of end user is secured in this research work by developing the attribute-based keyword search technique. Without knowing the true value of trapdoor and index, the cloud server performed the search with secure, which is ensured by constructing the secure ABE-based search technique. The proposed technique helped the numerous data owner for encrypting the data with various keys. The searching performance is improved and makes the process more natural by using the proposed ABE technique, where search request is completed by registered users without using the data owner's key. The trapdoor is generated, when the secret key is obtained by user from the CAs once the registration process is successful. The data are transmitted to cloud by user, where these data are encrypted by using re-encryption technique of management server, which will used to generate the trapdoor. The experiments are carried out on Wikipedia ontology to validate the performance of proposed ABE-based search technique in terms of various parameters such as precision, recall, searching time and efficiency in privacy preservation. When compared with existing techniques namely HABE, TFS, SII and so on, the proposed technique achieved higher recall (85.9%), precision (73.7%) and 99.2% privacy preservation efficiency with less searching time. However, the retrieval performance include precision is low, due to the length of the data. In future work, the length of the encrypted data can be reduced to increase the retrieval time of the searching keyword.

References

[1] H. Arshad, A. Jantan, G.K. Hoon, A.S. Butt, "A multilayered semantic framework for integrated forensic acquisition on social media," Digital Investigation, **29**, 147-158, 2019, doi:10.1016/j.diin.2019.04.002.

[2] A.H. Celdrán, M.G. Pérez, F.J.G. Clemente, G.M. Pérez, "Preserving patients' privacy in health scenarios through a multi context-aware system," Annals of Telecommunications, **72**(9-10), 577-587, 2017, DOI: 10.3233/THC-191731.

[3] R. Dubey, N. Namdeo, "Secure and Intelligent Decision Making in Semantic Web Mining using XML, XSLT and Xquery," International Journal of Scientific Research in Science, Engineering and Technology, **1**(6), 373-376, 2015.

[4] R. Bhatia, M. Singh, "Privacy Issues in Web Services: An Ontology Based Solution," Procedia Computer Science, **92**, 461-467, 2016, https://doi.org/10.1016/j.procs.2016.07.368.

[5] C.A. Ardagna, M. Cremonini,De S. Capitani di Vimercati, P. Samarati, "A privacy-aware access control system," Journal of Computer Security, **16**(4), 369-397, 2008, DOI: 10.3233/JCS-2008-0328.

[6] R. Bhatia, M. Singh, "An implementation model for privacy aware access control in web services environment," Proceedings of International Conference on ICT for Sustainable Development. Springer, Singapore, 2016, DOI: 10.1007/978-981-10-0129-1_50.

[7] H. Takabi, J.B.D Joshi, G. Ahn, "Security and privacy challenges in cloud computing environments," IEEE Security & Privacy, **8**(6), 24-31, 2010, DOI:10.1109/COMPSAC.2008.100.

[8] V.S. Thiyagarajan,A. Ayyasamy, "Privacy preserving over big data through VSSFA and MapReduce framework in cloud environment," Wireless Personal Communications, **97**(4), 6239-6263, 2017.

[9] G. Xu, Y. Cao, Y. Ren, X. Li, Z. Feng, "Network security situation awareness based on semantic ontology and user-defined rules for Internet of Things," IEEE Access, **5**, 21046-21056, 2017, DOI: 10.1109/ACCESS.2017.2734681.

[10] N.P. Nethravathi, P.G. Rao, V.J. Desai, P.D. Shenoy, K.R. Venugopal, M. Indiramma, "SWCTE: Semantic weighted context tagging engine for privacy preserving data mining", In 2016 International Conference on Data Science and Engineering (ICDSE), 1-5, 2016, DOI: 10.1109/ICDSE.2016.7823968.

[11] Y. Lu, O. S. Richard, "Semantic privacy-preserving framework for electronic health record linkage," Telematics and Informatics, **35**(4), 737-752, 2018, https://doi.org/10.1016/j.tele.2017.06.007.

[12] J. Liu, M. Zhou, L. Lin, H.J. Kim, J. Wang, "Rank web documents based on multi-domain ontology," Journal of Ambient Intelligence and Humanized Computing, 1-10, 2018, DOI:10.1007/S12652-017-0566-5.

[13] F.P. Appio, M.G. Cimino, A. Lazzeri, A. Martini, G. Vaglini, "Fostering distributed business logic in Open Collaborative Networks: an integrated approach based on semantic and swarm coordination," Information Systems Frontiers, **20**(3), 589-616, 2018, DOI: 10.1007/s10796-016-9691-5.

[14] Kalaivani, B. Ananthi, S. Sangeetha, "Enhanced hierarchical attribute based encryption with modular padding for improved public auditing in cloud computing using semantic ontology," Cluster Computing, 1-8, 2018, DOI:10.1007/s10586-018-2346-1.

[15] Valliyammai, A. Bhuvaneswari, "Semantics-based sensitive topic diffusion detection framework towards privacy aware online social networks," Cluster Computing, 1-16, 2018, DOI:10.1007/s10586-018-2142-y.

[16] Esposito, "Interoperable, dynamic and privacy-preserving access control for cloud data storage when integrating heterogeneous organizations," Journal of Network and Computer Applications, **108**, 124-136, 2018, https://doi.org/10.1016/j.jnca.2018.01.017.

[17] M. Imran-Daud, D. Sánchez, A. Viejo, "Privacy-driven access control in social networks by means of automatic semantic annotation," Computer Communications, **76**, 12-25, 2016, DOI:10.1016/j.comcom.2016.01.001.

[18] C. Ke, F. Xiao, Z. Huang, Y. Meng, Y. Cao, "Ontology-Based Privacy Data Chain Disclosure Discovery Method for Big Data," IEEE Transactions on Services Computing, 2019, DOI: 10.1109/TSC.2019.2921583.

[19] F. Boucenna, O. Nouali, S. Kechid, M. T. Kechadi, "Secure Inverted Index Based Search over Encrypted Cloud Data with User Access Rights Management," Journal of Computer Science and Technology, **34**(1), 133-154, 2019, DOI: 10.1007/s11390-019-1903-2.

[20] H. Tian, Y. Chen, C. C. Chang, H. Jiang, Y. Huang, Y. Chen, J. Liu, "Dynamic-hash-table based public auditing for secure cloud storage," IEEE Transactions on Services Computing, **10**(5), 701-714, 2017, DOI: 10.1109/TSC.2015.2512589.

[21] H. Hong, Z. Sun, X. Liu, "A key-insulated CP-ABE with key exposure accountability for secure data sharing in the cloud," KSII Transactions on Internet and Information Systems, **10**(5), 2394, 2016, DOI: 10.3837/tiis.2016.05.024

[22] J. Li, H. Wang, Y. Zhang, J. Shen, "Ciphertext-Policy Attribute-Based Encryption with Hidden Access Policy and Testing," Ksii Transactions on Internet & Information Systems, **10**(7), 2016, DOI: 10.3837/tiis.2016.07.026.

[23] H. Yin, L. Zhang, Y. Cui, "Improving Security in Ciphertext-Policy Attribute-Based Encryption with Hidden Access Policy and Testing," KSII Transactions on Internet & Information Systems, **13**(5), 2019, DOI: 10.3837/tiis.2019.05.029.

[24] T. Sauer, and Y. Xu, "On multivariate Lagrange interpolation," Mathematics of Computation, **64**(211), 1147-1170, 1995, DOI: https://doi.org/10.1090/S0025-5718-1995-1297477-5