METHODS OF HIDDEN PATTERN USAGE IN CLOUD COMPUTING SECURITY STRATEGIES WITH K- MEANS CLUSTERING

Swathi Priyadarshini
Research scholar, Department of computer science, Engineering college, Osmania University.
tigullaswathi@gmail.com

Dr. S. Ramachandram
Professor, Department of computer science engineering, Engineering college, Osmania University

.

Abstract:
Cloud computing has become one of the most important technologies for reducing cost and increasing productivity by efficiently using IT resources in various companies. The cloud computing system has mainly been built for private enterprise, but public institutions, such as governments and national institutes, also plans to introduce the system in India. In recent years, due to the appealing features of cloud computing, a large amount of data has been stored in the cloud. Although cloud-based services offer many advantages, the privacy and security of sensitive data is a big concern. It is desirable to outsource sensitive data in encrypted form to mitigate the problems. Encrypted storage protects the data against illegal access, but it complicates some essential functionality, such as the search on the data. A considerable amount of searchable encryption schemes has been proposed in the literature to achieve search over encrypted data without compromising privacy. Hover, almost all of them handle exact query matching but not similarity matching, a crucial requirement for real-world applications. Although some sophisticated secure multi-party computation-based cryptography techniques are available for similarity tests, they are computationally intensive and do not scale for significant data sources.

Keywords: *Data Security, Cloud Computing, Data Protection, Privacy, Risks and threats*

1.0 INTRODUCTION

Privacy of stored data is vital in many applications. Yet, it is becoming increasingly common for data to be hosted off-site, especially with the rise of cloud computing. Hover, cloud storage providers often cannot be trusted to respect the privacy of their host data, especially in the face of malicious insiders [1]. A simple solution is to encrypt the data before uploading it to the cloud. Hover, this would prevent the data from being searched. For example, users may wish to use their mobile phones to search their email. The cloud server will not identify which documents match the search query if the user's email data is encrypted using standard encryption techniques [2].

The cloud computing security area is classified into managerial, physical and technical area in the research, and then derives the detailed factors in each security area. The research derives the influence of security priorities in each area on the importance of security issues according to the identification of workers in private enterprise and public institutions [3]. Ordered prohibit models are used to analyze the influences and marginal effects of awareness for security importance in each area on the scale of security priority. The results show workers in public institutions regard the technical security as the highest importance, while physical and managerial security are considered as the critical security factors in private enterprise.

In practice, efficient unencrypted search algorithms usually use a precomputed database *index*. This allows keyword searches to be performed in essentially sublinear time concerning the size of the database (or, more precisely, in time proportional only to the number of *results* matching the query) [4]. Several index-based SSE protocols have been proposed, each more efficient than its predecessor.

2.0 Literature review

Song et al. [5] first proposed the concept of searchable symmetric encryption (SSE). As a new encryption primitive, Searchable encryption enables the user to search for a keyword over the ciphertext. Hover, the application was limited to search on static encrypted data and was unable to resist the simple adversary attack. Goh et al. [6] formally defined the secure index and developed a security model called the "semantic security" for adaptive selective keyword attacks. Hover, the accuracy of query result was limited due to the use of the Bloom filter. Curtmola et al. [7] proposed two new security models called "adaptive security" and "non-adaptive security", introducing a single key word-search SSE with a formal security definition. Due to the limitations of the SSE proposed earlier and the dilemma beten ensuring user privacy and efficient data usage on the cloud, Kamara et al. [8] introduced the dynamic searchable symmetric encryption (DSSE) method, which enabled the user to perform search and update operations on encrypted data Islam et al. [9] and Cash et al. [10] firstly exploited access pattern leakage and prior knowledge about the dataset to recover the user's query information. Liu et al. [11] exploited the search pattern to launch attacks and obtained users' query information. Zhang et al. [12] completely exposed the client's query and recovered user data and query information through the file injection attack. Simon et al. [13] leveraged both access and search pattern leakages to recover the keywords of queries. Therefore, an important direction for future research is to focus on the suppression of information disclosure, rather than setting it as default.

3.0 Research on Cloud Computing Security:

Security threats can easily occur to cloud computing services because IT resources are highly exposed to diverse attacks on computer networks due to the unique environment of the cloud computing services. For example, information for service subscribers is not safe from the attacks by hackers such as phishing and pharming. The information is stored in servers of service providers instead of the subscriber's computer. The hackers are able to easily obtain private information, business information, and computer or network information of subscribers because the information of all subscribers is stored and managed into the servers of the service providers. This work has expanded the research model that analyzed the importance of cloud computing security in three aspects. Similar to the previous research, the five-point Likert scale: not very important, not important, normal, important, very important, is used as a measurement standard of dependent variables, and the dependent variable has an ordered dummy value. When the dependent variable has an ordered number, the ordered choice model should be used instead of a general linear regression model. The ordered choice model is able to recursively deal with the survey responses investigated with the Likert scale. The ordered choice model is based on the normal choice model, which is used as the dependent variable, contains discrete data associated with choice. There are two types of ordered choice models as the ordered logit model and the ordered probit model.

Security Issues in Cloud Computing:

Security in the cloud is achieved, in part, through third party controls and assurance much like in traditional outsourcing arrangements. But since there is no common cloud computing security standard, there are additional challenges associated with this. Many cloud vendors implement their own proprietary standards and security technologies, and implement differing security models, which need to be evaluated on their own merits. In a vendor cloud model, it is ultimately down to adopting customer organizations to ensure that security in the cloud meets their own security polices through requirements gathering provider risk assessments, due diligence, and assurance activities Thus, the security challenges faced by organizations wishing to use cloud services are not radically different from those dependent on their own in-house managed enterprises. The same internal and external threats are present and require risk mitigation or risk acceptance. In the following, examine the information security challenges that adopting organizations will need to consider, either through assurance activities on the vendor or public cloud providers or directly, through designing and implementing security control in a privately owned cloud. In particular, examine the following issues:

- The treats against information assets residing in cloud computing environments.
- The types of attackers and their capability of attacking the cloud.
- The security risks associated with the cloud, and where relevant considerations of attacks and countermeasures.
- Emerging cloud security risks.
- Some example cloud security incidents.

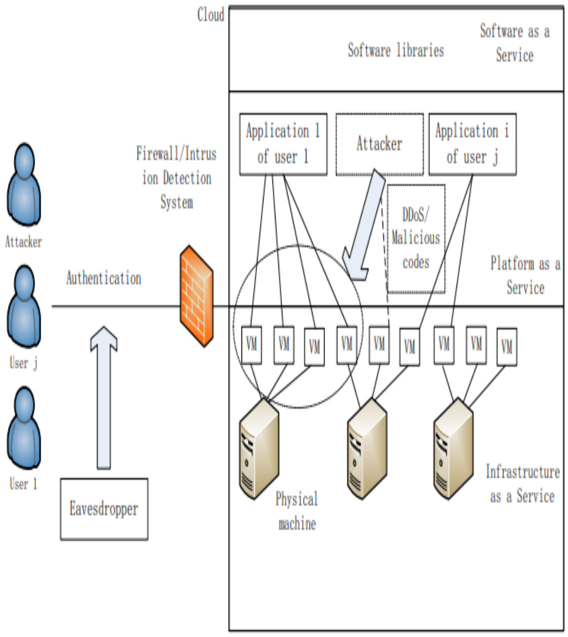
Cloud computing deployment model: Security challenges begin with the models for cloud service delivery. There are 4 types of models deployed in cloud computing

Public cloud: A public cloud depicts the conventional cloud computing where resources are dynamically monitored on a self-service basis over the Internet. This is done by implementing a third-party service provider that offers and share bills and resources via a registering utility basis. This cloud service focuses on a pay per usage model similar to the metering system for por and electricity, making it very flexible and adaptable, thus, attracting more demand for optimizing low security levels compared to other cloud models This is due to the extra effort in ensuring the security of all applications and information on the public cloud.

Private cloud: A private involves the use and control of cloud infrastructure in a personal/private cloud. It is mainly hosted in the central database of a firm and managed by either an internal person or a service provider. The advantage of this model are the individual controls the security of the system, Quality of Service (QoS), as ll as compliance.

Hybrid cloud: A hybrid cloud is a mixture of several cloud infrastructures (such as a private, public or community cloud infrastructure) which maintains the distinctive nature of its entity, but still connected through a standardized technology that allows data and application proprietary.

Community cloud: This is a collection of cloud infrastructure pooled by several administrators and provides support to communities with mutual concerns such as policy, consideration, security requirements, and compliance



Figure;1 Cloud security schematic environment

4.0 Hidden Vector Encryption and its Security:

Predicate encryption offers a new cryptographic mechanism that provides fine-grained access control over an encrypted database. In predicate encryption, decryption keys are associated with boolean predicates $f : \Sigma \longrightarrow \{0, 1\}$ over a pre-defined set of attributes Σ, while each ciphertext is associated with an attribute $I \in \Sigma$, and a payload messages $\mu \in M$. A decryption key can be used to decrypt a ciphertext only if the attribute I satisfies $f(I) = 1$. A major application of this encryption paradigm is to outsource encrypted data to a server, and yet retain the ability to make queries on the data without revealing more information than absolutely necessary. This is, in principle, similar to the concept of SSE described in the previous subsection. Hidden vector encryption (HVE) is one such predicate encryption scheme that supports conjunctive, equality, comparison, and subset queries on encrypted data. While HVE was formally defined in the public-key setting in [6], adopt their definition to the symmetric-key setting in order for it to be applicable in the context of SSE. A symmetric-key HVE may be defined as an ensemble of the following four PPT algorithms:

A symmetric-key HVE may be defined as an ensemble of the following four PPT algorithms:

- HVE. Setup(λ): takes a security parameter λ and outputs a master secret key msk. It also defines the message space M.
- HVE. Key Gen (msk, $v \in \Sigma$ m $*$): takes a predicate vector v, the master secret key msk and outputs a decryption key s.
- HVE. Enc (msk, $\mu \in M$, $x \in \Sigma$ m): takes as input a message μ, an index vector x, and the master secret keymsk and outputs the ciphertext c associated with (x, μ).
- HVE. Query (s, c): takes a ciphertext c corresponding to the index vector x and a decryption key s corresponding to the predicate vector v, and outputs the message μ if P HVE v (x) = 1.

say that a symmetric-key HVE is correct [6] if for all security parameters λ, all (μ, x) $\in M \times \Sigma$ m and all predicate vectors v, after sequentially running HVE. Setup(λ) to getmsk, HVE. Key Gen (msk, $v \in \Sigma$ m $*$) to get s, and HVE. Enc (msk, $\mu \in M$, $x \in \Sigma$

m) to get c, if P HVE v (x) = 1, then HVE. Query(s, c) = μ, otherwise

Pr [HVE. Query (s, c) =⊥] = 1 – negl (λ).

The next step is to formally define the notion of semantic security for symmetric-key HVE against PPT adversaries. The definitions are presented in the simulation-setting, which naturally subsumes the traditional security definitions for HVE in the indistinguishability setting. Prior to presenting the formal definition, present two auxiliary definitions that constitute the trivial leakage from any symmetric-key HVE scheme. Given a predicate vector v = (v1, . . . , vm) ∈ Σ m ∗ , its wildcard pattern α(v) is a vector of the same size as the predicate vector v, which is 1 if vj , ∗, and 0 otherwise. Also, given a predicate vector v ∈ Σ m ∗ and an index vector x ∈ Σ, the decryption pattern β (v, x) is a boolean value such that β(v, x) = 1 if P HVE v (x) = 1, and 0 otherwise. With these definitions in place, now define the real and simulation experiments for a symmetric-key HVE scheme.

4.1 Lightight symmetric-key hidden vector encryption

In this section, propose a novel HVE scheme in the symmetric key setting, referred to as SHVE, that entirely avoids the use of pairings. Our construction is predicate-only (implying that the payload message is "True" by default) and is amenable to parallel implementations for high performance. At the same time, it guarantees selective simulation-security against probabilistic polynomial-time adversaries for a single ciphertext query and an unbounded number of decryption key queries. Quite evidently, in our construction, the key-generation and query algorithms operate only on the secret-key/ciphertext components listed in the subset S, which correspond to the non-wildcard entries in the predicate vector. The speed-up achieved as a result of this property is particularly evident in applications where a majority of the predicate vectors have only sparsely distributed non-wildcard entries. As it turns out, our SSE construction, presented in the following section, presents precisely such an application scenario

4.2 Hidden pattern usage in cloud computing security strategies

Cloud computing and b services run on a network structure so they are open to network type attacks. One of these attacks is the distributed denial of service attacks. If a user could hijack a server, then the hacker could stop the b services from functioning and demand a ransom to put the services back online. To stop these attacks the use of syn cookies and limiting users connected to a server all help stop a DDOS attack. Another such attack is the man in the middle attack. If the secure sockets layer (SSL) is incorrectly configured then client and server authentication may not behave as expected therefore leading to man in the middle attacks. Another type of attack is network sniffing. With a packet sniffer an attacker can capture sensitive data if UN-encrypted such as passwords and other b service-related security configuration such as the UDDI (Universal Description Discovery and Integrity), SOAP (Simple Object Access Protocol) and WSDL (b Service Description Language) files. Port scanning is also another threat which can be used by an attacker. Port 80 is always open due to it being the port that the b server sits on. Hover, this can easily be encrypted and as long as the server software is configured correctly then there should be no intrusion

5.0 Results by using k-means

In brief, summarize our main contributions as follows:

1) In order to protect privacy information, we construct a new efficiency privacy-preserving *k*-means clustering scheme. In our setting, the scheme can both preserve the privacy of the data owners' sensitive information and the calculation results.

2) Existing works require the inputs to be encrypted under the same public key, which is very limited in practice. To avoid these problems, take advantage of proxy re-encryption to construct a scheme that is based on distributed data encrypted under multiple keys.

3) Utilize the two non-colluding servers model to complete large calculation in the learning phase. Except computing proxy keys, the data owners should do nothing during the learning process. In the end, the data owners only need to do decryption to obtain the result.

5.1 k-Means Clustering Algorithm

k-means clustering algorithm as one of the main data mining methods is widely used in practice. It can be used to partition a set of data objects into clusters. We assume that there are participants and each participant holds an -dimensional object ai(1≤k). *k*-means clustering algorithm aims to divide the objects into clusters Cj(1≤k), and to ensure a large degree of similarity within the same class, but little similarity between different classes. The clustering process is comprised of two steps. The first step is to assign objects into different clusters. The criterion of classification is to measure the distance between a sample Cj and the related cluster center Ci . There are many methods for this criterion, but in this paper, we adopt the Euclidean distance. At each iteration of the first step, *k*-means clustering algorithm assigns the object to the nearest cluster, which is labeled by , and it follows = argmin ||− ||2

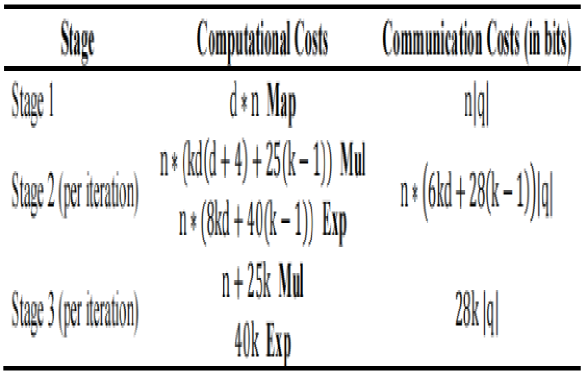
Assigning Records to the Nearest Cluster

The second step is to assign records to their nearest clusters by computing the minimum squared Euclidean distance. The cloud server S's first task is to initialize cluster centers .

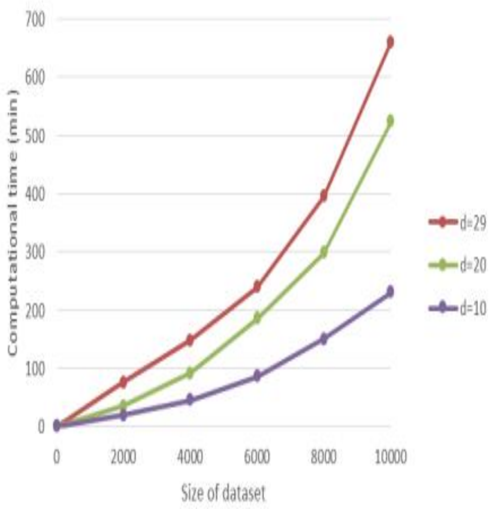
The general method is to initialize centers with randomly generated values. Alternatively, to reduce the number of iterations required in the process of clustering, we can adopt an optimized manner proposed in (Ostrovsky, Rabani, Schulman, & Swamy, 2012).

Assign the record to the cluster based on the minimum squared Euclidean distance and update the cluster label corresponding to the record at the same time. It is worth mentioning that the cluster label is encrypted with the cloud C's public key under the second-level encryption in our scheme. Its good point is the convenience for returning the final results to the data owners. Just like before, it only needs performing the proxy re-encryption to convert the ciphertexts to ones that encrypted under data owners' public keys.

5.2 Computational and communication costs of the scheme



Stage	Computational Costs	Communication Costs (in bits)
Stage 1	$d*n$ Map	$n\lvert q\rvert$
Stage 2 (per iteration)	$n*(kd(d+4)+25(k-1))$ Mul $n*(8kd+40(k-1))$ Exp	$n*(6kd+28(k-1))\lvert q\rvert$
Stage 3 (per iteration)	$n+25k$ Mul $40k$ Exp	$28k\lvert q\rvert$



Figure;2 Performance of the proposed scheme with varying size of datasets

6.0 Conclusions:

Cloud computing has been introduced to improve efficiencies of business and group productivity in private enterprises and public institutions. Current studies for security issues in cloud computing re on focused on addressing technical problems. Hover, to help establish a security strategy for introducing cloud computing systems, it is necessary to analyze security priorities and countermeasures. This paper discussed the risks and security threats to data in the cloud and given an overview of three types of security concerns. Virtualization is examined to find out the threats caused by the hypervisor. Similarly, threats caused by public cloud and multitenancy have been discussed. One of the major concerns of this paper was data security and its threats and solutions in cloud computing. Data in different states has been discussed along with the techniques which are efficient for encrypting the data in the cloud. The study provided an overview of block cipher, stream cipher and hash function which are used for encrypting the data in the cloud whether it is at rest or in transit. To meet the needs of practical application, we will continue improve the efficiency of the learning algorithm. Furthermore, we are planning to experiment with other machine learning algorithms used in other application scenario.

References:

1. *S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang and A. Ghalsasi, "Cloud Computing – The Business Perspective," Decision Support Systems, vol. 51, no. 1, pp. 176-189, 2011.*
2. *Korea Communications Commission and Korea Internet Security Agency, "Information Security guide for Cloud Services," Korea Communications Commissions and Korea Internet Security Agency, October, 2011*
3. *E. Y. Choi, B. J. Han, D. H. Shin, H. C. Jung and KISA Security R&D Team, "A Study for Enhancing Mobile Cloud Computing Security," in Proc. of 2011 Korean Society for Internet Information Summer Conference, vol. 12, no. 1, pp. 221-222, 2011*
4. *S. K. Eun, N. S. Cho, Y. H. Kim and D. S. Choi, "Cloud Computing Security Technology," Electronics and Telecommunications Trends, Electronics and Telecommunications Research Institute, vol. 24, no. 4, pp. 79-88, 2009.*
5. *C. S. Lim, "Cloud Computing Security Technology," Review of Korea Institutes of Information Security and Cryptology, vol. 19, no. 3, pp. 14-17, 2009*
6. *Song, D.X.,Wagner, D.,Perrig, A. Practical techniques for searches on encrypted data. In Proc. - IEEE Symp. Secur. Privacy. SP 2000. S&P 2000, pages 44–55. IEEE, 2000. 2. Goh, E.J. et al. Secure indexes. IACR Cryptol. ePrint Arch., 2003:216, 2003.*
7. *Curtmola, R.,Garay, J.,Kamara, S.Ostrovsky, R. Searchable symmetric encryption: Improved definitions and efficient constructions. 2006.*
8. *Kamara, S.,Papamanthou, C.,Roeder, T. Dynamic searchable symmetric encryption.*

In In Proc ACM Conf Computer Commun Secur, pages 965–976, 2012.

9. *Islam, M.S.,Kuzu, M.,Kantarcioglu, M. Access pattern disclosure on searchable encryption: ramification, attack and mitigation. In Ndss, volume 20, page 12. Citeseer, 2012.*
10. *Cash, D., Grubbs, P., Perry, J., Ristenpart, T. Leakageabuse attacks against searchable encryption. In In Proc ACM Conf Computer Commun Secur, pages 668–679, 2015.*
11. *Liu, C., Zhu, L., Wang, M., Tan, Y.A. Search pattern leakage in searchable encryption: Attacks and new construction. Inf Sci, 265:176–188, 2014.*
12. *Zhang, Y., Katz, J.,Papamanthou, C. All your queries are belong to us: The por of file-injection attacks on searchable encryption. In In Proc. USENIX Secur. Symp. (USENIX Security 16), pages 707–720, 2016*
13. *Oya, S., Kerschbaum, F. Hiding the access pattern is not enough: Exploiting search pattern leakage in searchable encryption. arXiv preprint arXiv:2010.03465, 2020.*