ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

Clustering and category used in Clinical Health Database

[1] P. SRINIVASA REDDY, [2]CHELLUBOINA SAI TEJA

[1](Assistant Professor), MCA, S.V.K.P & Dr K.S. Raju Arts & Science College

[2]MCA, scholar, S.V.K.P & Dr K.S. Raju Arts & Science College

ABSTRACT

Attribute-based encryption (ABE) offers a promising solution for flexible access control over sensitive personal health records in a mobile healthcare system on top of a public cloud infrastructure. However, ABE cannot be simply applied to lightweightdevices due to its substantial computation cost during decryption. This problem could be alleviated by delegating significant parts of the decryption operations to computationally powerful parties such as cloud servers, but the correctness of the delegated computation would be at stake. Thus, previous works enabled users to validate the partial decryption by employing a cryptographic commitment or message authentication code (MAC). This paper demonstrates that the previous commitment or MAC-based schemes cannot support verifiability in the presence of potentially malevolent cloud servers. We propose two concrete attacks on previous commitment or MAC-based schemes. We propose an effective countermeasure scheme for securing resource-limited mobile healthcare systems and provide a rigorous security proof in the standard model, demonstrating that the proposed scheme is secure against our attacks. The experimental analysis shows that the proposed scheme provides the similar performance compared with the previous commitment-based schemes and outperforms the MAC-based scheme.

1.INTRODUCTION

In mobile healthcare systems, medical devices are equipped with cyber capabilities and located close to patients to collect clinical data and report diagnostic information. Such devices could be semiconductor-embedded smart intelligent sensors which are implanted inside the patient's body and work for real-time quantification of pathological symptoms [1]. For diagnostic reports, personal health devices transfer private medical information to storage centers which manage these data in the form of electronic health record (EHR) [2]. Currently, many cloud service providers offer medical information services such as IBM Cloud Solutions for Healthcare [3], Google Cloud [4], and Azure for health [5] in practice. The cloud-based healthcare systems are beneficial since they can supply global connectivity to diverse healthcare devices and efficient resource maintenance in terms of storage saving. However, delegating sensitive information to the potentially untrusted cloud servers raises concerns with respect to the patient's privacy. Thus, fine-grained access control to the patients' data is of utmost importance to preserve privacy of the clients' sensitive medical information.

Attribute-based encryption (ABE) is a promising cryptographic tool to address this problem since it can provide pliable, fine-grained access control over data outsourced to the cloud [6], [7], [8]. Specifically, a data owner (a patient) generates an access policy, attaches to the encrypted message (a medical data), and transfers the ciphertext to the cloud server. As long as attributes assigned to a user (a medical practitioner) meet the policy enforced by the data owner, it is able to recover the message from the ciphertext1. Despite its promising aspect, unfortunately, ABE asks for considerable decryption operations on users. Moreover, such a cost exhibits a steady increase in linear growth as the access policy becomes bigger, i.e., more attributes and conditions are added to it.

Green et al. [9] gave a remedy to this issue by delegating to the cloud not only the ciphertext but also an ability to decrypt in partial on behalf of users. After the cloud partially decrypts and sends the result, a user decrypts the remaining part of the ciphertext of which computational cost is significantly smaller than to perform decryption from scratch. Thus, the great amount of computational burden on the user side can be offloaded to the cloud.

However, such a delegation of decryption raises a question about the correctness of the computation (even though it is partial) done by the cloud. In order to enable users to verify the partially decrypted ciphertext and attest to it, many studies adopted a commitment, a supplementary cryptographic tool that runs on top of ABE [10], [12]. In these schemes, a data owner generates a commitment value as well as a ciphertext which states that he commits to the encrypted message. Upon completing decryption, the user can use the commitment value to confirm that the partial decryption was done correctly. The commitment generation process, however, requires only public parameters, which implies that a malicious cloud is able to cheat users by forging a commitment for an arbitrary message and the original ciphertext.

To achieve resilience to the malicious attempt, there was an endeavour to replace the commitment with unforgeable message authentication codes (MAC) to provide the verifiability of the computation result [15]. Unfortunately, we observed that it is still vulnerable to the tampering attacks: an adversary can bypass the verification even without forging MAC (The detailed attack scenario will be given in Section 3.). Therefore, guaranteeing verifiable outsourced decryption of ciphertexts is an important step toward trustworthy delegation in a mobile healthcare system exploiting cloud computing.

Motivated by the above rationale, we propose a generic tamper-resistant commitment scheme for mobile healthcare cyber-physical systems in cloud. In the proposed scheme, a trusted authority issues a public verification key and a secret commitment key. Given a commitment key and a medical data to upload, a data owner generates a tamperresistant commitment value, encrypts the data, and uploadsthem to the cloud. After performing partial decryption, the cloud transfers the partially decrypted result to the user. He then performs final decryption, and uses the commitment value and the verification key to validate the correctness. Any malicious attempt to thwart the attestation results in the verification failure.

2.LITERATURE SURVEY

Clustering techniques have been massively used in the healthcare industry for easy diagnosis and prediction of diseases, thereby providing fast, adequate, reliable and less costly healthcare delivery to patients. Jabel and Srividhya [12], compared the performance of three clustering algorithms using heart dataset. They used Silhouette width measure to evaluate the performance of the algorithms, from their experimental results, CLARA clustering shows better performance compared to K-means, and PAM. The experiment was however limited to only partitioning clustering algorithms, ignoring other clustering algorithms such as Hierarchical and density-based clustering algorithms. In partitioning clustering, the number of clusters has to be specified by the developer, which can lead to incorrect clustering of the given dataset, while Hierarchical and Density-based clustering chooses the number of clusters by themself. Nithya et al. [20] extended International Journal of Computer Science & Information Technology (IJCSIT) Vol 10, No 2, April 2018 29 their research to include other clustering techniques. They compared the performances of three clustering algorithms-Hierarchical clustering, Density based clustering and k-Means clustering algorithms. Diabetes dataset was used to compare the performance of the algorithms based on their execution time and the number of clustered instances. The diabetic dataset was collected from UCI repository and it contains 769 instances and 9 attributes. They argued that using a training set parameter, k-Means algorithm gave better clustered instances compared to other clustering algorithms. They equally recommended other parameters such as cross validation, percentage split, and supplied test set. An exploratory data mining approach based on a densitybased clustering algorithm was also presented by [7]. For patients' clustering, they proposed a novel combine distance measure. Their aim was to discover cohesive and well-separated groups of diabetes patients with the same profile (i.e. age and gender) and examination history. They used diabetes dataset from an Italian Local Health Centre. Their experiment shows that their approach was effective in discovering groups of patients with the same examination history. Paul et al. [18], proposed K-means-Mode clustering algorithm using medical data. They pointed out that background knowledge of the medical domain in clustering process will increase the performance of the algorithm. They argued that their algorithm can handle both continuous and discrete data. Balasubbramanian and Umarari [2], analysed the effect of ground water on human health using clustering method. They used K-Means clustering algorithm to find out the risk factors related to the level of fluoride content in water. With the help of this analysis, they were able to discover useful hidden patterns which can help in the making of reasonable decisions in our society. Banu and Jamala [3], developed an algorithm for heart attack prediction using Fuzzy C means which is an unsupervised clustering algorithm that permits one data object to belong to more than one cluster. He argued that the proposed system will provide an aid for physicians to diagnose heart attack in a more efficient manner. Escudero et al. [9], in their work, used the concept of bioprofile and K-means clustering algorithm for early detection of Alzheimer disease and classification of Alzheimer disease into pathologic and non-pathologic groups. Zheng et al. [25] developed weight based K-means algorithm to identify the leukaemia, inflammatory, bacteria or viral infection, HIV infection and pernicious anaemia disease from hemogram blood test samples collected from Kovai Scan Centre. Their aim was to predict the diseases and also to find the efficiency of the algorithm. The performance of their algorithm was evaluated based on the clustering accuracy, execution time and error rate. They argued that their algorithm performed better than other clustering algorithms such as K-means and fuzzy c means. Belciug [5] used agglomerative hierarchical clustering algorithm to group patients according to their length of stay. This can help management in planning and decision making process. Alsayat and El-Sayed [1] presented an efficient K-means clustering algorithm that uses Self Organizing Map (SOM) method to overcome the challenge of finding number of centroids in ordinary k-means. They carried out performance evaluation of the algorithm using two healthcare datasets-Liver disease and heart disease datasets. They claimed that their proposed method shows better clustering performance. Belcliug et al. [6] assessed the effectiveness of three clustering algorithms using

Wisconsin Recurrence Breast Cancer dataset. They compared the performance of K-means algorithm with Self-Organizing Map and a cluster network implemented on a real-time decision support system for breast cancer recurrence detection. From their experiment, the result shows that cluster network had the highest performance. They argued that their clustering model showed a better diagnosing performance compared to the standard medical experience and it was also faster and cheaper. They pointed out that clustering method using imaging technique instead of database will likely perform better and will likely be more reliable. To bridge the gap between supervised and unsupervised learning, some researchers have combined clustering algorithms and classifier model for effective prediction and diagnosis. International Journal of Computer Science & Information Technology (IJCSIT) Vol 10, No 2, April 2018 30 Wisconsin Diagnostic Breast Cancer dataset from the University of California was also used by [26] to model algorithms for breast cancer diagnosis. They applied hybrid method of data mining using K-means and Support Vector Machine. K-means algorithm was used to uncover the hidden patterns of the Malignant and Benign tumours separately, while support vector machine was used to design a new classifier to detect breast cancer. Hybrid method of data mining was used by [4] to predict heart attack. They used K-Means to cluster the data after pre-processing. Maximal Frequent Itemset Algorithm (MAFIA) was used for mining maximal frequent patterns in the heart disease database. MAFIA is association rule mining techniques which are mainly efficient when the item set in the database is too large. The frequent pattern from the heart disease dataset was classified using C4.5 algorithm as the training algorithm. Applying the concept of information entropy, they argued that the designed prediction system is capable of predicting heart attack successfully. Ibrahim et al. [11] also applied hybrid the method of data mining. They compared the performance accuracy of Decision Tree Classifier with Agglomerative Hierarchical clustering to standard Decision Tree Classifier using diabetes dataset. The Agglomerative Hierarchical clustering algorithm was used to cluster the dataset, and then the resulting clusters were fed into the Decision Tree algorithm. They reported that the hybrid model achieved higher accuracy

3.SYSTEM ANALYSIS

Lai et al. [10] solved this issue by employing a checksum mechanism. However, their scheme is vulnerable to forgery attacks which enable an adversary to forge the checksum for any messages of his choice. Moreover, it requires encryption of both the messages and random values corresponding to checksums, which doubles the computational cost. To overcome this problem, Lin et al. [12] adopted a computationally lightweight commitment scheme to ABE. In the scheme, a sender commits to a message with a random coin and sends the result (i.e., commitment value) to a receiver. Provided the receiver recovers the message and the random coin correctly, he can verify whether the commitment value is valid. However, the sender generates a commitment value with only public parameters. Thus, Lin et al.'s scheme still suffers from the same forging attacks as Lai et al.'s scheme. Xu et al. [15] addressed this issue by adopting a MAC to each ciphertext. Since each MAC can be generated by only a sender who possesses a secret MAC key, an adversary can by no means forge the MAC without the key.

However, we observed that an adversary can bypass the verification procedure regardless of the MAC mechanism. In addition to verifiability in out sourceable decryption, several variations were introduced, aiming at providing extra capabilities such as outsourcing key issuing [11], key escrow [13], exculpability [14], outsourceable encryption [16], enhanced security (CCA security) [17], user revocation [18], and keyword search [19]. In this paper, we focus on verifiable outsourced decryption schemes based on commitment and MAC, and their security vulnerabilities. Specifically, we review Lin et al.'s scheme [12] and Xu et al.'s scheme [15] as representative commitment and MAC-based schemes, respectively, to show that they are all vulnerable to our attacks. However, it is important to note that our attack scenarios are not limited to them.

Disadvantages

❖ In the existing work, the system leaks significant information for updates and it is not parallelizable.

❖ The existing doesn't use MAC and very less security due to lack of ABE techniques.

3.1 PROPOSED SYSTEM

In the proposed scheme, a trusted authority issues a public verification key and a secret commitment keys. Given a commitment key and a medical data to upload, a data owner generates a tamper resistant commitment value, encrypts the data, and uploads them to the cloud. After performing partial decryption, the cloud transfers the partially decrypted result to the user. He then performs final decryption, and uses the commitment value and the verification key to validate the correctness. Any malicious attempt to thwart the attestation results in the verification failure. The contributions of this paper are:

The system proposes two attacks to reveal security breaches inherent in the existing verifiable outsourced decryption techniques. More precisely, we show that, in the commitment-based schemes, the cloud can bypass the verification by tampering with both the Cipher text and the corresponding commitment value. In the recently proposed MAC-based scheme, the system shows that the cloud can bypass the verification by forging the cipher text.

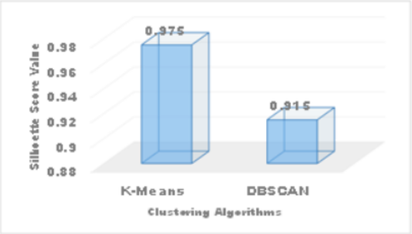
The system proposes a generic tamper-resistant commitment scheme for mobile healthcare cyber-physical systems in cloud, which can run on top of any ABE schemes with outsourced decryption capabilities. The system provides security and performance analyses to show that the proposed scheme provides tamper resistance for verifiable outsourced decryption, while rarely degrading efficiency compared with the existing schemes.

Advantages

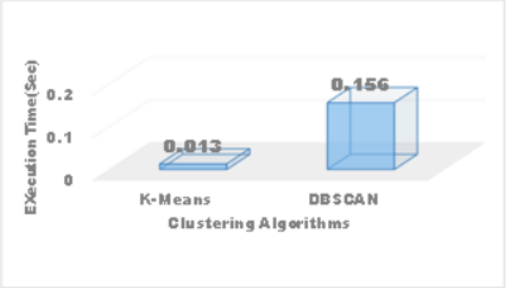
❖ Attribute-based encryption (ABE) is a promising cryptographic tool to address existing system problems since it can provide pliable, fine-grained access control over data outsourced to the cloud.

❖ The system is more secured since the MAC-based scheme tried to solve this problem by replacing the forgeable commitment with the unforgeable MAC, but we found that the malicious cloud can bypass the verification without tampering with MAC.

4. OUTPUTSCREENS



Comparison Of K-Means And DBSCAN In Terms Of Clustering Accuracy



Comparison of K-means and DBSCAN in terms of execution time

5. CONCLUSION

In this paper, we proposed two tampering attack scenarios to reveal the security breaches inherent in the existing verifiable outsourced decryption schemes. According to our analysis, in the commitment-based schemes, the cloud can skip the partial decryption by tampering with both the cipher text and the corresponding commitment value. Moreover, we showed that the cloud can bypass the verification in the MAC-based verifiable outsourced decryption scheme even though the MAC is unforgeable. We then proposed a generic tamper-resistant commitment scheme for mobile healthcare cyber-physical systems in cloud. The proposed scheme can run on top of any ABE schemes with outsourced decryption capabilities. We provided security and performance analyses to show that the proposed scheme provides tamper resistance while rarely degrading efficiency compared with the existing schemes.

6. REFERENCE

[1] Lee, I., Sokolsky, O., Chen, S., Hatcliff, J., Jee, E., Kim, B., and Venkatasubramanian, K. K., "Challenges and research directions in medical cyber–physical systems," Proceedings of the IEEE, 100(1), pp. 75–90, 2012.

[2] Jovanov E., O'Donnell Lords A., Raskovic D., Cox P. G., Adhami R., and Andrasik F., "Stress monitoring using a distributed wireless intelligent sensor system," IEEE Engineering in Medicine and Biology Magazine, vol. 22, no. 3, pp. 49–55, 2003.

[3]https://www.ibm.com/cloud/healthcare.

[4https://cloud.google.com/solutions/healthcare-life-sciences.

[5] https://azure.microsoft.com/en-us/industries/healthcare.

[6] Goyal, V., Pandey, O., Sahai, A., and Waters, B. “Attribute-based
encryption for fine-grained access control of encrypted data”, In
Proceedings of the 13th ACM conference on Computer and communications
security, pp. 89–98, 2006.

[7] Bethencourt, J., Sahai, A., and Waters, B. “Ciphertext-policy
attribute-based encryption”, In Security and Privacy, 2007. SP’07.
IEEE Symposium on, pp. 321–334, IEEE.

[8] Ostrovsky, R., Sahai, A., andWaters, B. “Attribute-based encryption
with non-monotonic access structures”, In Proceedings of the 14th
ACM conference on Computer and communications security, pp. 195–
203, ACM, 2007.

[9] Green, M., Hohenberger, S., and Waters, B. “Outsourcing the
decryption of abeciphertexts”, In USENIX Security Symposium, Vol.
2011, No. 3, 2011.

[10] Lai, J., Deng, R. H., Guan, C., andWeng,J.“Attribute-based encryptionwith verifiable outsourced decryption”, IEEE Transactions on Information Forensics and Security, 8(8), pp. 1343–1354, 2013.