Ensuring Privacy of Medical Data in Cloudlet-Based Storage System

1Dr.Manoj Challa 2Shreya C Bhatt 3Usha N 4 Vidhur RK
CMR Institute of Technology, Bengaluru.

Abstract: Cloud storage is increasingly being used by enterprises due to the on-demand scalability and pay as you go models offered by cloud providers. Medical companies outsource patient medical data to the cloud. It is very important to ensure the safety of the data stored in the cloud, as data loss can compromise the safety of the patient. This task proposes a secure and privacy-protecting data storage mechanism for cloudlet-based storage. The patient data collected from wearable sensors are split to shares using Chinese remainder algorithm and distributed to cloudlets. Without compromising a minimum number of cloudlets, it becomes difficult for the attacker to reconstruct the data. By this way, the security and privacy for patient health care data is ensured.
Keywords: Cloud Storage, Scalability, Security, Cloudlets, Health Care

I. INTRODUCTION

Body area networks(BAN) are used to provide elderly health care. The sensors are attached to patient body and the data collected from these sensors are sent to cloudlets-based storage systems. The data is monitored to provide timely health care to the patients. The first application of BAN is expected primarily in the health care sector, especially in the ongoing monitoring and differentiation of critically ill patient with chronic diseases such as diabetes, asthma and heart diseases, and in the monitoring of elderly care. Other new applications for this technology include military, fire-fighting, security, gaming,
social computing, entertainment and sports. The
large amount of data collected by the BAN node should be stored in the cloud. This requires storage / processing, a strong and secure infrastructure.
The potential integration of BANs and cloud computing introduces a viable hybrid platform that needs to be able to handle large amounts of data collected from multiple BANs. BAN relays useful and important information to the cloud that can operate in a distributed and hostile environment, so new security mechanisms are needed to prevent malicious interactions with the storage infrastructure. Both the cloud providers and the users must take strong security measures to protect the storage infrastructure. This work addresses the problem and propose a secure and privacy preserving algorithm for secure distribution of health care data to cloud.

II. LITERATURE SURVEY

A survey of secure and privacy preserving algorithms for cloud storage is discussed in this section.
Wei wang et al [1] have proposed a hybrid cloud based solution to ensure privacy. In this solution, insensitive data is kept in public cloud and sensitive data is sent to private cloud. Two protocols are proposed in this work, for providing privacy and defending against collusion between the cloud providers and users. The protocols were validated for bounded data distortion. The proposed solution was tested against real work datasets. Through experimental analysis, the solution is found to be scalable for large datasets. The solution has certain disadvantages for health care data as it does not have the capability to support queries specific to health care data. Xueli Huang and Xiaojiang Du [2] have proposed a hybrid cloud based privacy preservation scheme. The method

has comparatively lower storage, computation and communication overhead. The method can work both for text and image data. Through experimental analysis, the proposed method is found to have better data security of AES algorithm. But the problem in this solution is – it distributes more load to private cloud than public cloud. As the result, there is more capital expenditure on enterprises in setting up higher capacity private cloud. Also, computation overhead is higher for image data. Hui Zhang et al [3] have proposed a cost-effective cloud in hybrid mode. It makes the best use of legacy data center in the enterprise for private cloud and integrates it with public cloud services. The application workload is segregated into base workload and flash workload and distribution of load between private and public cloud is done based on work load type. A FFDI detection algorithm is proposed to categorize the work load based on multiple criteria like incoming requests, volume of data handled, changes in application priority etc. Simulation analysis showed that the system is able to achieve higher resource utilization of private and public clouds for the workloads. Jin Li et al [4] proposed a novel encryption scheme for cloud outsourced data. The method is a de-duplication scheme which can ensure higher level of confidentiality for sensitive data. Differential access for user over the de-duplicated data is implemented in this proposed solution. A proof of concept of the proposed solution is implemented in the live cloud and the solution is found to have lower overhead. But the problem with this solution is that, complexity increases linearly with data dimensionality. Thus the method cannot be applied for health care data as they generally have multiple dimensions. Jingwei Li et al [5] proposed hybrid architecture with private and public cloud for addressing the privacy preserving without compromising on data utility. Private cloud acts as middle interface between user and public cloud. Both access control at a finer level and keyword based retrieval is supported in this work. The proposed solution can be easily extended with fuzzy keyword search. Xuyun Zhang et al [6] proposed a privacy preservation quasi-identifier index addressed privacy preservation without compromise on data utility. The approach can work for incremental dataset. Experimental results demonstrate the effectiveness of the proposed solution over other for incremental dataset. The problem in this approach is that it computation wide complex due to involvement of too many hashes. Kui Ren et al [7] ensured the integrity of privacy preserved data in cloud using auditing. A trusted third-party agent audits the data in the cloud. The auditing is done without any compromise to the privacy of the data as it does not need any decryption. The solution is found to be computationally efficient through experimental analysis. The solution becomes inefficient when specialization or generalization is considered and for data involving frequent modification. H. Liu et al [8] proposed a authentication protocol with privacy preservation for cloud storage. The solution is designed considering three factors of security aspects like access control. Access control was implemented using attribute-based encryption. For enhanced data sharing Proxy re-encryption is applied. The solution is found to be effective for multi-user collaboration in cloud. The disadvantage in this solution is its lack of scalability for large datasets. The solution is centralized and suffers from single point of failure.S. Rani et all in [9] proposed a technique for sharing of health care data between organizations in a secure and privacy preserving manner. The solution is built on the assumption of semi trusted clouds. Cryptographic secret sharing is applied for sharing the data. Distribution of data across multiple clouds is done using cryptographic secret sharing. The problem in this solution is that it is not very efficient for data retrieval. Stefanos et all in [10] proposed a scalable data anonymization technique. Privacy breaches are handled using the semantic relationship between the sensitive attributes. The records in the data are clustered using agglomerative clustering. Each cluster is encrypted using differential privacy algorithm.

This solution too is not efficient for retrieval. Zahoor A. Khan in [11] enforced security of the health care data using attribute based encryption. Access control at fine level is enforced using attribute based encryption. Authors experimented with different versions of attribute based encryption schemes and found that the use of CP-ABE has comparatively lower computation cost. A. Champinget all in [12] proposed a solution to secure health care data to cloud using a modified attribute-based encryption. The modified attribute-based encryption is able to solve the key distribution issues in the cloud. The method is also able to simplify the key distribution process using implicit authentication. A major problem in this solution is that it is centralized and suffers from single point of failure. J.Yang et all in [14] proposed a data sharing solution for cloud. A hybrid solution has four key concepts: to separate data vertically before publishing; data-based access; data validity testing and mixed mathematical search and cryptography are four concepts used in this work to use data effectively without compromising privacy protection. Kao et all in [15] proposed a data perturbation using reversible privacy contrast mapping (RPCM) algorithm. The algorithm has two steps for distraction and recovery. By combining nearby values, distortion is created. Watermarking is done in addition to perturbation. Watermarking ensures integrity of data.

III. PROPOSED METHODOLOGY

Figure 2 shows the architecture of the proposed methodology. The system consists of BAN records. BANs consist of multiple users (each user has a BAN installed), who cannot send the data collected by BAN without the main unit. A group of BAN users can be integrated around a single cloudlet server. This represents a small amount of cloud computing capacity enough to manage BAN users in the cluster. The most important part of a cloudlet server is the storage system. Storage systems need to provide measurable and reliable storage space for large amounts of data. Various cloudlet systems can be connected to each other via wireless or wireless communication links (such as WiMax).In addition, the cloudlet system can connect directly to the business cloud system using wireless or wireless communication links. A business cloud system is a central management and storage area that can be accessed by various organizations that are interested in a particular type of data. After the data has been collected from the patients' attached machines, the data is encrypted using a numerical research unit method. Data is transferred to the nearest cloudlet. The data in the cloudlet can also be stored in the cloud, based on user preferences but in this case, the data is segmented, so that privacy is not leaked. The module process flow diagram is given in Figure 1.

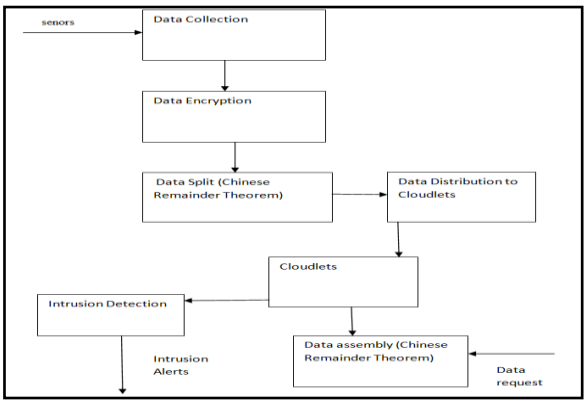

Figure 1 Process flow

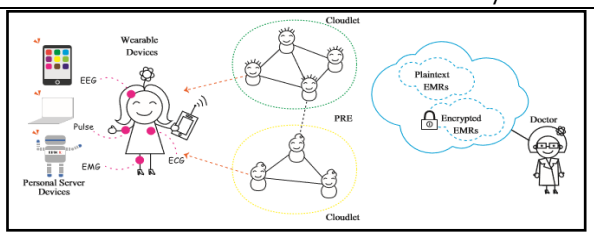

Figure 2 System architecture

The data collected from the sensor is encrypted using AES (Advanced Encryption Standard) algorithm. The encrypted data is split to N shares using Chinese remainder algorithm. The shares are split in such a way that reconstruction of data can be done with M (M<N) unique shares. Each of the cloudlet are distributed with M-1 shares. By this way shares from minimum two cloudlets are needed to reconstruct the data.

When there is a need to retrieve the data, the users send the retrieval request to cloudlet manager. Cloudlet manger places the request to active cloudlets. Cloudlets send the corresponding shares to the Cloudlet manager. Cloudlet manager reconstruct the data, decrypts the data and forwards to the requested user.

Security is enforced in this work in following ways.

The connection between cloudlet manger and cloudlets is fully secure. The data is distributed to cloudlet only after mutual authentication between cloudlet manager and cloudlets. Only a minimal of M-1 shares are distributed to cloudlet, due to this attacker has to minimally compromise two cloudlets. Also, a curious cloudlet cannot reconstruct the data by itself. For additional security, the data is encrypted and splited to share. Users requesting the data are authenticated by the cloudlet manager before providing the requested data to the users.

IV. RESULTS

The performance of the proposed solution is measured in terms

1. Data storage time
2. Data retrieval time

Data storage time involves the time for encryption using AES, time for split of shares using Chinese remainder algorithm and time for distribution of shares to cloudlets.

Data retrieval time involves the time for retrieving the shares from the cloudlets, time for reconstruction from shares and time for decryption using AES.

The data storage and retrieval time is measured for different volumes of data and compared with time taken for storage and retrieval without any protection on cloudlets. The results for data storage time for different size of file is given below in table 1.

Table1: Data Storage for different file sizes

File size(KB)	Proposed (milli sec)	Plain storage to cloudlet(milli sec)
1	13	10
10	101	93
100	958	876

1000	7854	6125

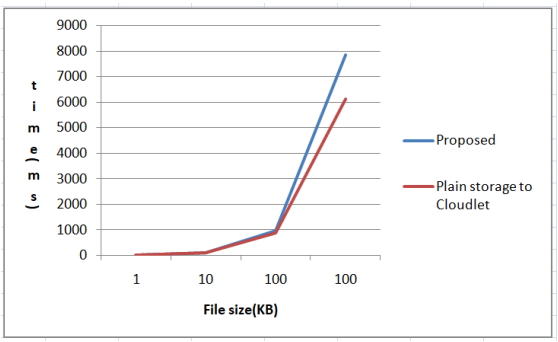

Figure 3: Graphical representation of data storage

When the size of file has been increased the average data storage time has also increased by 25.61% in proposed solution compared to Plain storage to cloudlet as shown in figure 3. This extra time is justified for the security and privacy protection offered by the proposed solution.

The results for data retrieval time for different size of file is given in the below table 2.

Table 2: Dataretrieval for different file sizes

File size(KB)	Proposed (milli sec)	Plain storage to cloudlet (milli sec)
1	7	5
10	72	51
100	678	523
1000	6124	5623

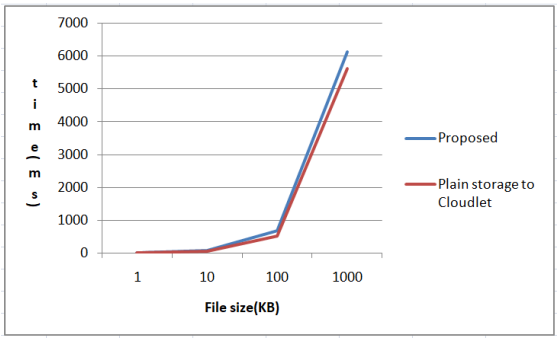

Figure 4: Graphical representation of data retrieval

During the data retrieval when the size of the file increases the average retrieval time also has increased by 10.96% in the proposed solution compared to plain storage to cloudlet as shown in figure 4. The overhead is due to reconstruction using Chinese remainder algorithm and AES decryption. This is only 11% overhead for implementing security and privacy which is justifiable.

V. CONCLUSION

In this work, we proposed a secure and privacy preserving algorithm using Chinese remainder algorithm for data storage in cloudlets. Through three means of encryption with AES, authentication of all connections, data split and distribution stronger security and privacy is offered by the proposed solution. The solution is able to provide this enhanced protection at cost of 25% increment in data storage time and 10% increment in data retrieval time. Exploring other means of split to reduce the data storage and retrieval overhead is in scope of future work.

REFERENCES

[1] W. Wang, L. Chen and Q. Zhang, "Outsourcing high-dimensional healthcare data to cloud with personalized privacy preservation", Computer Networks, vol. 88, pp. 136-148, 2015.

[2] X. Huang and X. Du, "Achieving data privacy on hybrid cloud", Security and Communication Networks, vol. 8, no. 18, pp. 3771-3781, 2015.

[3] H. Zhang, G. Jiang, K. Yoshihira and H. Chen, "Proactive Workload Management in Hybrid Cloud Computing", IEEE Transactions on Network and Service Management, vol. 11, no. 1, pp. 90-100, 2014.

[4] J. Li, Y. Li, X. Chen, P. Lee and W. Lou, "A Hybrid Cloud Approach for Secure Authorized Deduplication", IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 5, pp. 1206-1216, 2015.

[5] J. Li, J. Li, X. Chen, Z. Liu and C. Jia, "Privacy-preserving data utilization in hybrid clouds", Future Generation Computer Systems, vol. 30, pp. 98-106, 2014.

[6] C. Wang, S. Chow, Q. Wang, K. Ren and W. Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage", IEEE Transactions on Computers, vol. 62, no. 2, pp. 362-375, 2013.

[7] X. Zhang, C. Liu, S. Nepal and J. Chen, "An efficient quasi-identifier index based approach for privacy preservation over incremental data sets on cloud", Journal of Computer and System Sciences, vol. 79, no. 5, pp. 542-555, 2013.

[8] H. Liu, H. Ning, Q. Xiong and L. Yang, "Shared Authority Based Privacy-Preserving Authentication Protocol in Cloud Computing", IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 1, pp. 241-251, 2015.

[9] S. Rani, J. Malhotra, R. Talwar, "EEICCP-Energy Efficient Protocol for Wireless Sensor Networks", Wireless Sensor Network, vol. 5, no. 7, pp. 127-136, 2013.

[10] Stefanos A. Nikolidakis, Dimitrios D. Vergados, Christos Douligeris Algorithms, Energy Efficient Routing in Wireless Sensor Networks Through Balanced Clustering, 2013.

[11] Zahoor A. Khan, Shyamala Sivakumara, William Phillips, Bill Robertson, "A QOS-aware Routing Protocols for Reliability Sensitive Data in Hospital Body Area Networks", Trans. on ELSEVIER in proc. ANT, pp. 171-179, 2013.

[12] Achampong, Emmanuel & Dzidonu, Clement. (2016). Optimising Attribute-based Encryption to Secure Electronic Health Records System within a Cloud Computing Environment. 27-34. 10.21742/ijcs.2016.3.2.04.

[13] Jin Sun, Xiaojing Wang, "A searchable personal health records framework with fine-grained access control in cloud-fog computing",PLOS ONE,2018

[14] J. Yang, J. Li, and Y. Niu, "A hybrid solution for privacy preserving medical data sharing in the cloud environment", Future Generation Computer Systems, Vol. 43-44, No. 2, pp. 7486, 2015.

[15] Kao, Yuan-Hung & Lee, Wei-Bin & Hsu, Tien-Yu & Lin, Chen-Yi & Tsai, Hui-Fang & Chen, Tung-Shou. (2015). Data Perturbation Method Based on Contrast Mapping for Reversible Privacy-preserving Data Mining. Journal of Medical and Biological Engineering. 35. 10.1007/s40846-015-0088-6.