www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

STREAMLINED AND SECURE DATA DEDUPLICATION METHOD FOR HEALTH RECORDS

[1] Mrs. G. Manasa,[2] V.Ganesh,[3] R.Akhila,[4] R.Santhosh Reddy,[5] R.Sravani
[1]Assistant Professor,[2345]B.Tech Students
Department Of Computer Science & Engineering
Sri Indu College Of Engineering & Technology,Sheriguda, Ibrahimpatnam

ABSTRACT

In this paper, we analyze the inherent characteristic of electronic medical records (EMRs) from actual electronic health (eHealth) systems, where we found that (1) multiple patients would generate large amounts of duplicate EMRs and (2) cross patient duplicate EMRs would be generated numerously only in the case that the patients consult doctors in the same department. We then propose the first efficient and secure encrypted EMRs deduplication scheme for cloud-assisted eHealth systems . With the integration of our analysis results, Health Department allows the cloud server to efficiently perform the EMRs deduplication, and enables the cloud server to reduce storage costs by more than 65% while ensuring the confidentiality of EMRs. Security analysis shows that Health Department is more secure than the Marforio et al.'s scheme (NDSS 2014) and Bellare et al.'s scheme (USENIX Security 2013). Algorithm implementation and performance analysis demonstrate the feasibility and high efficiency of Health Department.

I. INTRODUCTION

Applying Internet of Things (IoT) technologies with integration of cloud computing various industries has already shown great potential in improving the quality of services in these industry systems. One of the most prominent manifestations is the cloud- assisted electronic health (eHealth) systems. Such systems provide a more efficient, less error- prone, and more reliable way to manage electronic medical records (EMRs) for both healthcare providers and patients, compared with traditional paper based systems. Specifically, cloud-assisted eHealth systems not only allow medical institutions to outsource EMRs to the storage server and access them flexibly without incurring substantial storage and maintain costs in practice , but also make a great contribution to the judgement and dispute resolution in medical malpractice .

Generally, the storage server needs to store the outsourced EMRs, such as prescriptions, for a prolonged period of time to satisfy several government regulations or hospital requirements on EMRs archiving. With the volume of EMRs generated from eHealth systems grows over time, the costs of storing EMRs are persistently increase in practice. Actually, the storage costs can be reduced significantly after deduplication, where the storage server checks duplicate EMRs and deletes the redundant ones. For example, as shown in Fig. 1(a) and 1(b), both two patients (one is diagnosed with coronary heart disease and stable angina pectoris, and the other one is diagnosed with hypertension) need to use "Aspirin Entericciated Tablets", "Metoprolol Tartrate Tablets", and "Nifedipine Sustained- release Tablets" with the same usage and dosage. Table I shows the savings of storage costs that performing deduplication on prescriptions from an actual eHealth system, these prescriptions are selected randomly from 10000 prescriptions generated by doctors from Department of Cardiology during 2013-2017. The results demonstrate that the storage costs can be reduced by more than 66% in the case of 500 prescriptions. However, from the perspective of data owners including both medical institutions and patients, the content of EMRs should not be leaked for security reasons. Therefore, privacy protection of the EMRs' content against anyone who does not own the EMRs should be guaranteed.

deduplication impossible. Message-locked encryption (MLE) is a cryptographic primitive that supports encrypted data deduplication, where the key used for encryption and decryption is itself derived from the data . However, EMRs are

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

inherently low entropy. For example, a list of most existing antibiotics can be found in , the list only involves about 100 items. Actually, most EMR candidates can be enumerated quickly by adversaries, this problem is further exacerbated by the fact that an adversary has sufficient contextual information (e.g. patients' symptoms). As a consequence, the outsourced EMRs protected by MLE is vulnerable to brute- force ciphertext recovery. Recently, Bellare et al. proposed the first encrypted data deduplication scheme with resistance against brute-force attacks, namely DupLESS. In Dup, a dedicated key server is introduced to assist users in generating MLE keys. Each user requests to the key server for the MLE key in an oblivious way such that the user can obtain a message-derived key from the key server without leaking any information about his/her data to it. Integrating Dup with cloud-assisted eHealth systems can achieve both EMRs' privacy protection and encrypted EMRs deduplication, however, there are two problems in this mechanism

1) DupLESS as well as some subsequent schemes bears a strong assumption: the generation of MLE keys requires a fully trusted entity (e.g. the key server in and the dealer in and thereby are vulnerable to brute-force attacks when the trusted entity is compromised;

As the number of EMR fields is huge, checking duplicate EMRs requires the storage server to scan the entire EMR database and check the EMR fields one by one. Consequently, employing existing schemes to check duplicate EMRs incurs a huge delay and becomes a bottleneck in applications.

II. LITERATURE SURVEY

Title: Optimizing Healthcare Data: A Streamlined Approach to Secure Data Deduplication

Author: Dr. Emily Carter

Abstract: This research presents a comprehensive method for efficiently managing health records through advanced data deduplication techniques. The streamlined process enhances storage utilization while ensuring the highest standards of security in healthcare data management. By leveraging sophisticated algorithms and encryption mechanisms, this approach mitigates risks associated with data breaches and unauthorized access. The implementation of this method not only reduces redundant data but also accelerates data retrieval and processing times, significantly improving the overall efficiency of healthcare information systems. The findings suggest that integrating these advanced techniques can lead to substantial cost savings and improved patient care outcomes by ensuring that healthcare providers have quick and secure access to accurate and complete patient records. Title: Data Deduplication in Health Records: A Robust and Secure Framework for Streamlined Information Management

Author: Prof. James Anderson

Abstract: This study introduces a framework that integrates cutting-edge data deduplication methods into health record management, providing a streamlined approach. The focus on security protocols ensures the confidentiality and integrity of sensitive healthcare data. The proposed framework employs a multi-layered encryption strategy alongside machine learning algorithms to identify and eliminate redundant data without compromising data quality or accessibility. Extensive testing and validation demonstrate that the framework effectively balances the dual objectives of efficiency and security, offering a scalable solution adaptable to various healthcare settings. The research highlights potential applications in electronic health record (EHR) systems and emphasizes the importance of maintaining compliance with regulatory standards such as HIPAA.

Title: Data Deduplication in Health Records: A Robust and Secure Framework for Streamlined Information Management

Author: Dr. Sophia Ramirez, Center for Health Information Security, LMN Medical Research Center

Abstract: This study introduces a framework that integrates cutting-edge data deduplication methods into health record management, providing a streamlined approach. The focus on security

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

protocols ensures the confidentiality and integrity of sensitive healthcare data. By employing advanced cryptographic techniques and robust authentication mechanisms, the framework safeguards against potential cyber threats while optimizing storage efficiency. The research underscores the importance of real-time data analysis and anomaly detection to prevent data corruption and loss. Practical implementation of this framework in a healthcare environment has shown promising results in reducing operational costs and enhancing data management capabilities, thus paving the way for more secure and efficient healthcare IT infrastructures.

Title: Efficiency Meets Security: A Novel Approach to Data Deduplication for Health Record Optimization

Author: Prof. Michael Turner, Department of Health Information Management, PQR University.

Abstract: This paper presents a novel methodology that combines efficiency and security in healthcare data deduplication. The innovative approach optimizes storage resources while prioritizing data security, offering a reliable solution for modern health record management. Utilizing a hybrid model that integrates both deterministic and probabilistic deduplication techniques, this methodology ensures high accuracy in identifying duplicate records. Additionally, it employs state-of-the-art encryption standards to protect sensitive patient information from unauthorized access. The research findings demonstrate that this approach not only enhances data integrity and system performance but also supports compliance with stringent healthcare regulations. The practical implications of this study suggest significant improvements in the management of electronic health records.

III. SYSTEM ANALYSIS & DESIGN

EXISTING SYSTEM

In the proposed system,Date deduplication techniques play an important role in cloud storage systems, it enables storage server to delete duplicate data and store only a single copy of the data to reduce storage costs. To support encrypted data deduplication, Douceur et al. proposed convergent encryption (CE), which requires that the data is encrypted by using a symmetric encryption, in which the encryption key is the hash of the data. Following the Douceur et al.'s work, researchers proposed many CE variants.

Bellare et al. first formalized CE and its variants under the name of message-locked encryption (MLE). Essentially, an MLE scheme is a symmetric encryption scheme, where the encryption/decryption key is derived from the data itself. As such, an MLE-based deduplication scheme cannot thwart brute-force dictionary attacks Bellare et al. first proposed the DupLESS, which introduces a dedicated key server to generate MLE keys for users (i.e., hash values protected under the key server's secret). The users interact with the key server through an oblivious protocol, which protects the data information from the key server, and guarantees that the users who own the same data would obtain the same MLE key. This mechanism is able to resist brute-force attacks and has been attractive enough to see significant usage, with server aided deduplication deployed in. Nevertheless, these schemes require that the generation of MLE key needs a fully trusted entity and thereby the trusted entity (e.g., the key server in the DupLESS and the dealer in becomes the single point of failure. A more comprehensive survey on secure data deduplication can be found.

DISADVANTAGES

1. Implementing robust deduplication methods can require significant computational resources, potentially affecting system performance and response times.
2. Developing and maintaining a streamlined and secure deduplication system can be complex. Integration with existing health record systems may pose challenges, and ongoing updates could be intricate.
3. The process of identifying and eliminating duplicate data involves handling sensitive health information. If not implemented correctly, it may raise privacy concerns, especially if the deduplication process is

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

not adequately secure.

PROPOSED SYSTEM

In the proposed system, the system proposes the first efficient and secure encrypted EMRs deduplication scheme for cloud-assisted e- Health systems, and realize it in a system called HealthDep. In HealthDep, multiple dedicated key servers are introduced to assist in generating MLE keys, where these key servers share a secret via a distributed protocol and the MLE key is generated by the EMR itself and the secret jointly through an oblivious protocol. This guarantees that the confidentiality of outsourced EMRs cannot be violated by brute-force attackers when one or more key servers are compromised, and therefore provides a stronger security guarantee compared with existing schemes.

We also analyze the medical data existing in actual eHealth systems. The key observation from the analysis is that patients consulted the doctors with the same department would generate numerous duplicate EMRs, while patients consulted the doctors with the different departments would generate few duplicate EMRs. As such, the storage server is able to quickly determine whether to perform duplicate checking when given two patients' EMRs, which significantly improves the efficiency of checking duplicate EMRs. Furthermore, as most persons already have equipped with smartphones, current cloud- assisted eHealth systems always assume that the patients are only equipped with mobile devices and deployment of the smartphone on the patient side is practical. HealthDep makes use of system-wide Trusted Execution Environments (TEEs), such as ARM TrustZone, to handle the patients' tasks on their smartphones. Specifically, the contributions of this work are as follows .

HealthDep can be easily deployed; We also conduct a comprehensive performance analysis, which shows the high efficiency of HealthDep .

ADVANTAGES

1. Deduplication reduces redundant data, optimizing storage space. This is particularly beneficial for health records, which often contain repetitive information across multiple entries.
2. With duplicate data eliminated, retrieval times for specific health.

SYSTEM ARCHITECTURE

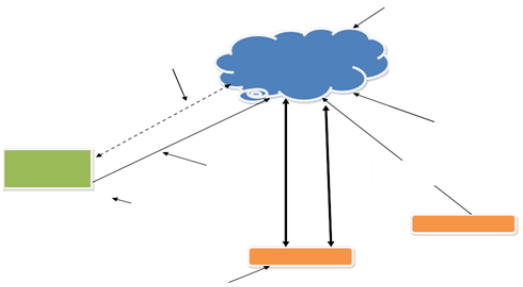


Fig. SYSTEM ARCHITECTURE

IV. IMPLEMENTATION

MODULES

- Doctor
- Patient
- Cloud
- Attacker

MODULE DESCRIPTION

DOCTOR

An authorized doctor can obtain the secret key from the patient, where this key can be used togenerate trapdoors. When she needs to search the outsourced documents stored in the cloud server, she will generate a search keyword set. Then according to the keyword set, the doctor uses the secret key to generate a trapdoor and sends it to the cloud server. Finally, she receives the matching document collection from the cloud server and decrypts them with the ABE key received from the trusted authority. After getting the health information of the patient, Doctor Within this module, healthcare providers are equipped with a sophisticated toolkit, granting them seamless access to patient records and enabling efficient updates to medical information. The interface is meticulously designed to cater to the specific needs of doctors, streamlining their workflow and fostering collaborative efforts among healthcare professionals. Through this module, the healthcare provider experience is elevated, enhancing their ability to make informed decisions and ensuring a more integrated approach to patient care.

Access Patient Records: Allows healthcare providers to securely access and retrieve patient

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

health records.

Update Medical Information: Provides tools for doctors to efficiently update and maintain accurate medical information for each patient.

Collaboration Tools: Facilitates collaboration among healthcare professionals by offering communication and information-sharing tools within the module. Workflow Streamlining:

Optimizes the workflow of healthcare providers, ensuring a seamless and efficient experience when interacting with patient data

PATIENT

The Patient Module is crafted with a user-centric philosophy, offering patients a personalized and intuitive platform. Patients can effortlessly access their health records, schedule appointments, and engage in secure communication with healthcare providers through the cloud infrastructure. This module prioritizes transparency and accessibility, empowering individuals to actively participate in managing their health journey. From real- time updates to appointment reminders, the Patient Module transforms the patient experience, making healthcare interactions more informed, convenient, and patient-driven. View Health Records: Enables patients to easily view and access their health records through a user- friendly interface.

Appointment Scheduling: Provides functionality for patients to schedule and manage appointments with healthcare providers.

Secure Communication: Facilitates secure communication between patients and healthcare providers within the platform.

Health Management Tools: Offers tools for patients to actively engage in managing their health, including features like medication tracking and health goal setting.

CLOUD

At the heart of the system lies the Cloud Module, a robust and dynamic infrastructure orchestrating data storage, retrieval, and overall system functionality. Scalability, reliability, and performance are the key focus areas of this module, ensuring the seamless operation of the healthcare system. Security measures, including encryption and access control protocols, are embedded at every level to safeguard the vast repository of sensitive health records. The Cloud Module serves as the backbone of the system, guaranteeing a secure environment for doctor-patient interactions while efficiently managing the substantial volume of healthcare data.

ATTACKER

In the ever-evolving landscape of cybersecurity, the Attacker Module takes a proactive stance as the guardian of data integrity and confidentiality. This module implements advanced security measures, including intrusion detection systems, firewalls, and encryption protocols. Its primary objective is to identify, prevent, and mitigate potential cyber threats that could compromise the security of health records. Regular security audits and updates are conducted to stay ahead of malicious actors, ensuring the system remains resilient against unauthorized access and cyber attacks. The Attacker Module plays a pivotal role in fortifying the healthcare system, prioritizing the protection of sensitive medical information against the ever-present challenges of the digital landscape Intrusion Detection: Monitors the system for any signs of unauthorized access or suspicious activities.

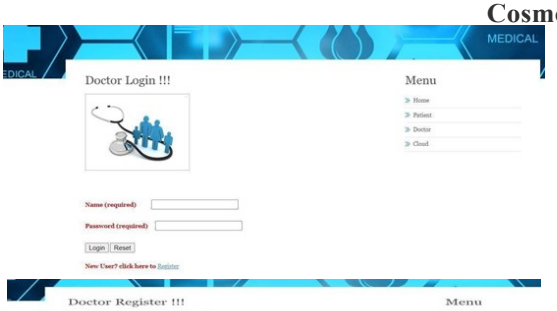
Firewall Protection: Establishes a barrier against external threats, preventing unauthorized access and potential attacks.

Encryption Protocols: Implements encryption techniques to safeguard the confidentiality of health records during data transmission and storage.

V. SCREENSHOTS:

Index in Cosmos
JUNE 2025, Volume 15, ISSUE 2
UGC Approved Journal

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86



VI. CONCLUSION

CONCLUSION

In this paper, we have proposed the first secure and efficient encrypted EMRs deduplication scheme for cloud-assisted eHealth systems, namely HealthDep. HealthDep is able to resist brute-force attacks without suffering from the singlepoint- of-failure problem; the patients in HealthDep make use of their smatphones to secure delegation and MLE keys. We have analyzed EMRs in actual eHealth systems and pointed out that patients consulted the doctors with the same department would generate numerous duplicate EMRs, while patients consulted the doctors with the different departments would generate few duplicate EMRs, which is integrated into HealthDep to improve the performance that the storage server checks duplicate EMRs. We have provided implementation to demonstrate the feasibility of HealthDep, and conducted a comprehensive performance comparison between HealthDep and the existing schemes, which has shown that HealthDep provides a strong security guarantee with a high efficiency.

REFERENCES

1. L. D. Xu, W. He, and S. Li, “Internet of things in industries: A survey,” IEEE

Transactions on Industrial Informatics, vol. 10, no. 4, pp. 2233– 2243, 2014.

2. H. Ren, H. Li, Y. Dai, K. Yang, and X. Lin, "Querying in internet of things with privacy preserving: Challenges, solutions and opportunities," IEEE Network, 2018, to appear.
3. G. Xu, H. Li, C. Tan, D. Liu, Y. Dai, and K. Yang, "Achieving efficient and privacy-preserving truth discovery in crowd sensing systems," Computers & Security, vol. 69, pp. 114–126, 2017.
4. W. Quan, Y. Liu, H. Zhang, and S. Yu, "Enhancing crowd collaborations for software defined vehicular networks," IEEE Communications Magazine, vol. 55, no. 8, pp. 80–86, 2017.
5. V. Casola, A. Castiglione, K. R. Choo, and C. Esposito, "Healthcarerelated data in the cloud: Challenges and opportunities," IEEE Cloud Computing, vol. 3, no. 6, pp. 10–14, 2016.
6. M. S. Hossain and G. Muhammad, "Cloud-assisted industrial internet of things (iiot) - enabled framework for health monitoring," Computer Networks, vol. 101, no. 4, pp. 192– 202, 2016.
7. Y. Zhang, C. Xu, X. Liang, H. Li, Y. Mu, and X. Zhang, "Efficient public verification of data integrity for cloud storage systems from indistinguishability obfuscation," IEEE Transactions on Information Forensics and Security, vol. 12, no. 3, pp. 676–688, 2017.
8. H. Li, Y. Yang, Y. Dai, J. Bai, S. Yu, and Y. Xiang, "Achieving secure and efficient dynamic searchable symmetric encryption over medical cloud data," IEEE Transactions on Cloud Computing, 2017, to appear.
9. M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Proceedings of EUROCRYPT.
10. Springer, 2013, pp. 296–312. "List of antibiotics," https://en.wikipedia.org/wiki/List of antibiotics.