Security of Medical Big Data Images using Decoy Technique

US. Bhargavi
Department of CSE, Manipal Institute of Technology, Manipal Academy of Higher Education
Manipal, India
bhargavisreevathsa@gmail.com

Shivaprasad Gundibail
Department of CSE, Manipal Institute of Technology, Manipal Academy of Higher Education
Manipal, India
shiva.prasad@manipal.edu

KN. Manjunath
Department of CSE, Manipal Institute of Technology, Manipal Academy of Higher Education
Manipal, India
kn.manjunath@ieee.org

A. Renuka
Department of CSE, Manipal Institute of Technology, Manipal Academy of Higher Education
Manipal, India
renuka.prabhu@manipal.edu

***Abstract*—Tele-radiology is a technology that helps in bringing the communication between the radiologist, patients and healthcare units situated at distant places. This involves exchange of medical centric data. The medical data may be stored as Electronic Health Records (EHR). These EHRs contain X-Rays, CT scans, MRI reports. Hundreds of scans across multiple radiology centers lead to medical big data (MBD). Healthcare Cloud can be used to handle MBD. Since lack of security to EHRs can cause havoc in medical IT, healthcare cloud must be secure. It should ensure secure sharing and storage of EHRs. This paper proposes the application of decoy technique to provide security to EHRs. The EHRs have the risk of internal attacks and external intrusion. This work addresses and handles internal attacks. It also involves study on honey-pots and intrusion detection techniques. Further it identifies the possibility of an intrusion and alerts the administrator. Also the details of intrusions are logged.**

Keywords—Medical Big Data, Cloud Computing, Security, Decoy, Honey-pot, Intrusion Detection.

I. INTRODUCTION

The idea of telemedicine which emerged in early 1960's brought a new vision to healthcare sector. It is extremely useful in cases where patient's transport from one hospital to another is almost impossible. The history and origin of idea of telemedicine and implementing a nation-wide telemedicine network is discussed by Jader Wallauer, Douglas Macedo, Rafael Andrade and Aldo von Wangenhelm [1]. The doctor and patient can do a video conference. The doctors situated at distant places can discuss simultaneously on critical cases. Therefore, many prototypical models for telemedicine networking came into picture. Across the decades by the evolution in technology, these prototypes have become working models and are used worldwide. Electronic Health Records (EHRs) help in diagnosis and analysis. On the other hand, these EHRs are so voluminous and form Big Data. These data are termed as Medical Big Data (MBD). Therefore, a new cloud computing paradigm emerged to provide Healthcare As A Service (HAAS) [2][3]. According to Mario F. Li and Jun Feng [2], apart from providing help in analysis and diagnosis, a healthcare cloud can be a platform for idea exchange, knowledge base and communication in aspects of medicine. However even healthcare cloud is susceptible to security threats as other clouds. General threats to a cloud as discussed by DIAO Zhe, WANG Qinghong, SU Naizheng and ZHANG Yuhan [4] include attacks during Data Transfer, attacks to Data Storage and Cloud Terminal Disruption. D. Sangita, C. Ankita and P. Reshamlal [5] mention that the security issues also include legal and policy related concepts, software licensing etc. The concept of security involves two ideas in major. One is keeping the things secure i.e. away from attacks and disruption. Other is detecting an intruder and study of intrusions to prevent further attacks. G. Snehal Kene and P. Deepti Theng [6] have proposed a survey on existing Intrusion Detection Systems in the domain of cloud computing. An intruder may be an unauthorized user trying to view or manipulate data or an authorized user misusing privileges provided. One of the technique used in intrusion detection is to deviate an intruder by providing false data or a compromised data. The hub which contains the compromised data is referred as a Honeypot. According to Lance Spitzner [7] the definition of a honeypot is, a resource which can be probed or attacked and compromised. Sometimes data compromising is achieved by creating a decoy of the existing data. In a general sense decoy can be described as, to entice a person into a trap thus taking him/her away from the intention held.

The major challenge of honeypot technology is to secure the data in spite of compromising the data. Also to study the behavior and profiling of intruders to prevent the attacks. This challenge served as a motivation for take up of this work.

The paper is organized as follows. In Section II the literature review is discussed. This covers a detailed study of healthcare cloud, security threats and Intrusion Detection Systems. Section III describes design and methodology of an intrusion detection system for MBD using decoy data in a shared environment like cloud. The results, analysis are discussed in section IV. Section V concludes the work.

II. LITERATURE REVIEW

Technology has evolved contemporary world to such an extent that the communication establishment between individuals across the countries is a matter of not more than few minutes. Extending the technological features to every aspect including agriculture and medical aid has proven its strength to sophisticate human life.

According to Mario F.Li and Jun Feng [2] though development in technology has changed the social

communication perspectives to a broader range through facebook or twitter or any such general platforms, there is no agreeable platform for healthcare society to communicate among itself. The healthcare society involves healthcare individuals like doctors, nurses, pharmacists and pharmaceutical organizations, medical infrastructure providers, device manufacturers, patients and patient's care takers. Having a shared platform among healthcare individuals helps in diagnosing, analyzing and hopefully solving patient's problems. Cloud computing paradigm can serve a better service in this regard. A healthcare cloud can be built on the network infrastructure. Many public and private cloud providers such as Amazon, Microsoft and Google provide large storage with extremely fast computational capability. On such a cloud, healthcare professionals can share their ideas in analyzing a case or an unsolved case. They can put their knowledge and views on emerging technological devices and their benefits in medical IT. A paper by Hanen Jemal, Zied Kechaou, Mounir Ben Ayed and Adel M.Alimi [3] states that the concept of healthcare cloud remarkably helps in Biomedical Informatics. The writers also mention about the recent reports on cloud usage by healthcare organizations. According to the statistics, in the year 2015 about 30% of the healthcare organizations were using cloud service for their needs.

With the increase in usage of cloud services the threats and attacks have also increased correspondingly. The web page at [8] has discussed the most recent types of attacks happened to public cloud. These types of attacks include breaches in data, an enterprise placing an unsecure critical software on cloud thus giving a potential chance for an attack, unencrypted data being attacked, improper authentication, threat from an insider or an authorized user, hijacking of the account and DDoS attacks. Amit Hendre and Karuna Pande Joshi [9] have discussed possible threats and some of the control models which can be implemented for cloud security. They state that along with data breaches, denial of service and account hijacking there exist few more threats like Insecure interfaces and APIs, service traffic hijacking, abuse of cloud services, exploiting shared technology vulnerabilities.

Many Compliance Standards or Control Models have been discussed by National Institute of Standards and Technology (NIST), Cloud Security Alliance (CSA) [10][11][12]. Some of them include standards for

- *Encryption of Data and Key Management:* Use FIPS [13] or Vaultive [14]
- *Multi Media Protection:* Use MPAA [15]
- *Authentication, Authorization and Identification:* Use STIG [16] or FedRAMP [17]
- *To achieve Virtualization and maintaining the Resource Abstraction:* Use DMTF-CADF [18], PCI-DSS [19].

Any breach to the security for an intended misuse can be termed as Intrusion. William Stallings states in the book [20] that an intruder can be a Masquerader (Unauthorized user exploiting the legitimate user's account) or a Misfeasor (Authorized user misusing the privileges) or can be a Clandestine user (Stealing supervisory access and evading the auditing information). Whenever the best system of security to prevent intrusion fails, the system's next hope will be an Intrusion Detection System (IDS). Intrusion Detection can help in 2 ways:

1) Strong IDSs can detect the intrusion so quickly that the intruder can be denied access before data gets compromised or damaged.
2) The pattern of the intrusion recorded can serve to prevent the further attacks.

IDSs notify the administrator and also members of team (if permission given by administrator) about the intrusion. According to Manish Kumar and M. Hanumanthappa [21] IDSs can detect intrusion through either of the following two basic mechanisms:

1) Signature based detection: The Patterns or the signatures of previous attacks commonly termed as Known Attacks are recorded. Upon these previous records an intrusion can, be suspected.

2) Anomaly based detection: Any deviation from the expected normal behavior is termed as an anomaly. In most of the cases user profiling will be done and predictive algorithms will be used.

IDS involve a technique to deviate the attacker from the attack. This can be achieved through honeypot. The definition of honeypot has been stated in introduction. K. G. Anagnostakis et al. [22] have discussed how honeypots can be used to detect unkown attacks. The unknown attacks will not be having any signature or pattern available. Hence IDSs create honeypot data where a portion of data get compromised at the cost of identifying an intruder, his/her intentions and pattern of attack. This section focuses on some of the properties of honeypot. Wenjun Fan, Zhihui Du, David Fernandez and A. Victor Villagra [23] have proposed an anatomic view of honeypot systems. Accordingly, a honeypot contains two important components namely Decoy and Captor. Decoy can be viewed as the false data or compromised data susceptible to attacks. Captor defines the actions to be taken when decoy gets disturbed i.e., possibly an intrusion had taken place. Authors also present a model of honeypot diagrammatically as in Fig. 1.

The compromised data have been kept in a decoy file. According to the degree of compromised data, honeypots can be classified as Low Interaction Honeypots (LIH), Medium Interaction Honey-pots (MIH), High Interaction Honeypots (HIH) [23]. The desirable properties of a decoy file are discussed by Al Hamid et al. [24]. Authors state that a decoy file should be believable, should be capable of enticing and should be differentiable from original file.

The important challenge of decoy file lies in implementing the differentiability property. To provide enough enticing through honeypot and detecting the intrusion is the objective of the work.

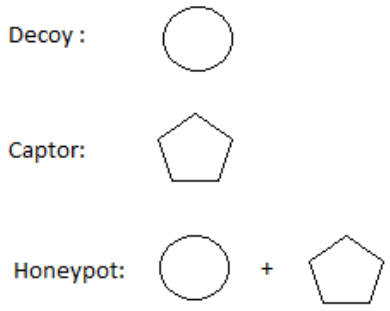

Fig. 1. Model of Honeypot

III. METHODOLOGY

The proposed work focuses on detecting an intruder by intentionally compromising a portion of the data. These data are termed as decoy data. A decoy image is constructed for every original medical image stored in the cloud database. The design of the work is as shown in Fig. 2.

Decoy database is kept as a honey-pot and continuous monitoring is done on the honey-pot. Any update like an insertion of an image file into the database or modifying and updating the contents of the existing image or deletion of a record in the database is an anomaly.

The technique implemented is more of an anomaly based intrusion detection rather than signature based. An anomaly based intrusion detection helps in detecting the unknown attacks. The next level of authentication is needed to extend the security. In this step, the identity of the individual with respect to healthcare IT can also be verified.

Usually images are stored as blob objects or entities in the cloud. Blob refers to Binary Large Objects. Blob objects are used for image or video storage. Following services are provided by Blob storage [25]:

- Massive, unstructured data can be stored.
- Images and documents can be sent to the browser directly.
- Distributed access for files can be achieved.
- Audio or Video streaming can be served.
- Large log files can be created and maintained.
- Storage of data for the purpose of backup, disaster recovery or archiving can be achieved.

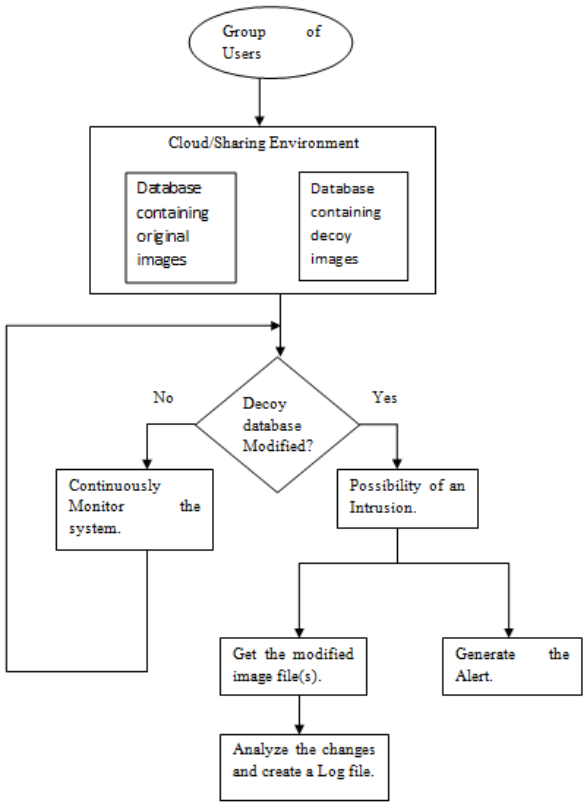

Fig. 2. Intrusion Detection System for Medical Big Data in Cloud/Shared environment.

Objects in the blob storage can be accessed by a client program through simple HTTP and HTTPS. Some of the programming languages like PHP, Python, Ruby, Java, .NET and Node.js include storage client libraries to help storage access of blob objects.

The proposed work uses DICOM images as input data. DICOM is a standard for medical image communication. Most of the processing such as creation of decoy, image analysis is done at a local machine. Then, the data are uploaded to cloud. Cloud is mainly used for storing the large database of decoy and original images and sharing of access to users.

IV. RESULTS AND ANALYSIS

The proposed technique being focused on Anomaly based Intrusion Detection, recognizes any modification to the decoy file as an attempt for intrusion. Whenever decoy data are disturbed, the system sends an alert to administrator's personal device or mailbox. Accordingly, administrator can take further actions. If the administrator is interested to analyze the pattern of attack, access to decoy for intruder can be continued. Otherwise, access for the shared data can be denied immediately.

The DICOM images taken from Cancer Imaging Archive data set [26][27] were used in the proposed research work. One of the very important challenge of the decoy is that the decoy should be believable. An original image and its corresponding decoy are shown in Fig. 3 and Fig. 4.

To check the similarity between constructed decoy image and its corresponding original image, a metric known as Structural Similarity Index (SSIM) is considered. The SSIM value is calculated for the collection of original images and their corresponding decoy images for the above data set. For the collection of data set being considered, the value of SSIM for each pair of original and its decoy is in the range of 0.90 to 0.99. The maximum value of SSIM is 1 for exactly same images. Hence, it can be inferred that a high similarity is achieved between the original and the decoy image. Higher the value of SSIM, lower is the human perception for dissimilarity. Hence a high similarity ratio achieved will entice the intruder. The intruder starts believing that he/she has got the original data. This is a method of taking the intruder away from his/her intention.

Fig. 3. Original dicom image under consideration.

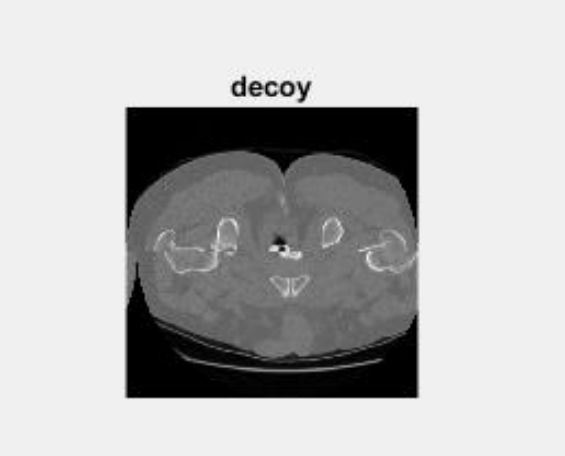

Fig. 4. Corresponding decoy image for Fig. 3.

Each and every modification done to decoy database is notified to the system administrator. To study the intrusion pattern in detail, the modified image(s) are taken and compared against their respective decoy images. Mean Square Error (MSE) is calculated to identify the extent of modification. For manual alteration of the decoy chosen, with 25 pixels in random, the MSE value is noted to be 22.6449. Here, the MSE value depends on the value replaced with the original pixel value, Section of the image considered for modification etc. The results are stored in a log file. This log file can potentially serve as an input for intrusion pattern detection and behavioral analysis of the intruder.

The proposed work has been tested on public cloud and it has given the expected results. The types of the notifications sent to the administrator under different circumstances is listed in the TABLE I.

TABLE I. EVENTS AND NOTIFICATIONS

Name of Folder on Cloud	*Access Holders*	*Event*	*Notification Type*
Root	Administrator	Creation/Update of File(s)	General
Original images	Administrator	Creation of File(s)	General
Original images	Administrator	Update of File(s)	Alert Message
Decoy images	Administrator, Users	Creation/Update/Deletion of File(s)	Alert Message

V. CONCLUSION

The need for security has been increased to a greater extent since the emergence of cloud paradigm. A shared platform like cloud would be more vulnerable to attacks compared to a local machine. When sensitive data like medical images have to be placed and managed on a cloud platform, a number of security measures must be taken care of. The known and unknown attacks can be identified, analyzed and also prevented to some extent using Intrusion Detection Systems. But the implementation of the IDS must be so efficient that the time taken to detect the intrusion should be less than the time taken to entice the intruder. The proposed technique, mainly focused on identifying the intruder, may be a normal user from public community or an authorized user misusing the privileges. The intentions of the attack and working principles of the attack can also be predicted to some extent. The work can be extended to take appropriate actions against intrusions with still minimum degree of compromised data. The computations can be shifted to fog nodes so that the speed can be achieved along with the security.

ACKNOWLEDGMENT

We thank, Dr. Clark, Cancer Imaging Archive, USA, who permitted us for using their DICOM images in the research.

REFERENCES

[1] Jader Wallauer, Douglas Macedo, Rafael Andrade, and Aldo von Wangenheim, “Building a national telemedicine network”, IT professional 10.2, Mar, 2008.

[2] Mario F. Li., & Jun Feng, “Healthcare road map to modernization in clouds: healthcare forum for healthcare professionals, medical device manufacturers, pharmaceutical companies and average people on virtual private clouds”, In Proceedings of the Second IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies, pp. 247-248, IEEE Press, July 2017.

[3] Jemal, Hanen, Zied Kechaou, Mounir Ben Ayed, and Adel M. Alimi. "Cloud computing and mobile devices based system for healthcare application." In *Technology and Society (ISTAS), 2015 IEEE International Symposium on*, pp. 1-5. IEEE, 2015.

[4] Zhe, Diao, Wang Qinghong, Su Naizheng, and Zhang Yuhan. "Study on Data Security Policy Based on Cloud Storage." In *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2017 IEEE 3rd International Conference on*, pp. 145-149. IEEE, 2017.

[5] D. Sangita, C. Ankita and P. Reshamlal, “A review on issues and challenges of cloud computing”, Int.J. Innov.Adv.Comput.Sci., vol 4, no. 1, pp. 81-88, 2015.

[6] G. Snehal Kene., and P. Deepti Theng., "A review on intrusion detection techniques for cloud computing and security challenges.", In *Electronics and Communication Systems (ICECS), 2015 2nd International Conference on*, pp. 227-232. IEEE, 2015.

[7] Lance Spitzner, Honeypots : Tracking Hackers, Boston, United States: Addison Wesley, pp. 052-060, 2002.

[8] Alex Bennett, “*8 Public Cloud Security Threats to Enterprises in 2018*”, Apr 2018. [Online]. Available : http://www.comparethecloud.net/articles/8-public-cloud-security-threats-to-enterprises-in-2017/?cn-reloaded=1. [Accessed: 08-Oct-2018].

[9] Amit Hendre and Karuna Pande Joshi, "A semantic approach to cloud security and compliance", In *Cloud Computing (CLOUD), 2015 IEEE 8th International Conference on*, pp. 1081-1084. IEEE, 2015.

[10] NIST Recommended Security Controls for Federal Information Systems and Organizations, http://csrc.nist.gov/publications/nistpubs/800-53-Rev3/sp800-53-rev3-final_updated-errata_05-01-2010.pdf

[11] CSA, Diana Kelley, “Understanding Cloud Controls Matrix v1.4.xls”.

[12] CSA, Security Guidance Version 3, 14/11/2011.

[13] FIPS 140-2, May 2001, "Security requirement for cryptographic models”, http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf.

[14] Vaultive, http://www.vaultive.com/technology/encryption-in-use.

[15] MPAA, Entertainment content security and protection, http://fightfilmtheft.org/docs/2012_Annual_Trending_Report_Final.pdf, 2012.

[16] STIG, Application Security and Development STIG, 2014.

[17] FedRAMP, https://www.gsa.gov/technology/government-it-initiatives/fedramp/about-fedramp.

[18] DMTF CADF, June 2012, Cloud Auditing Data Federation http://www.dmtf.org/sites/default/files/standards/documents/DSP2028_1.0.0a.pdf .

[19] PCI-DSS , Oct 2010, “Requirement and security assessment”, https://www.pcisecuritystandards.org/documents/pci_dss_v2.pdf

[20] William Stallings, “Cryptography and Network Security”, 5th Ed, United States : Pearson, pp. 748-755, 2002.

[21] Manish Kumar, and M. Hanumanthappa. "Scalable intrusion detection systems log analysis using cloud computing infrastructure." In Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on, pp. 1-4. IEEE, 2013.

[22] K. G. Anagnostakis, S. Sidiroglou, P. Akritidis, K. Xinidis, E. Markatos, and A. D. Keromytis, “Detecting targeted attacks using shadow honeypots,” in Proc. 14th Conf. Usenix Security Symp., Berkeley, CA, USA, vol. 14, pp. 9–25, 2005.

[23] Wenjun Fan, Zhihui Du, David Fernández, and Victor A. Villagra. "Enabling an Anatomic View to Investigate Honeypot Systems: A Survey.", IEEE Systems Journal, 12, pp. 3906-3919, 2017.

[24] Al Hamid, Hadeal Abdulaziz, Sk Md Mizanur Rahman, M. Shamim Hossain, Ahmad Almogren, and Atif Alamri, "A security model for preserving the privacy of medical big data in a healthcare cloud using a fog computing facility with pairing-based cryptography”, IEEE Access, 5, pp. 22313-22328, 2017.

[25] Introduction to object storage in Azure”, Oct 2018. [Online]. Available:https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blobs-introduction. [Accessed: 22-Oct-2018].

[26] National Cancer Institute (NCI), www.cancerimagingarchive.net (2016), https://public.cancerimagingarchive.net/ncia/login.jsf. Accessed : 27th Feb 2016.

[27] Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P. et al, ”The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository”, Journal of Digital Imaging, 26(6), pp. 1045-1057, 2013.