35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Steganography-based Access Control to Medical Data Hidden in Electrocardiogram

Vu Mai, Ibrahim Khalil, Ayman Ibaida

Abstract

Steganographic techniques allow secret data to be embedded inside another host data such as an image or a text file without significant changes to the quality of the host data. In this research, we demonstrate how steganography can be used as the main mechanism to build an access control model that gives data owners complete control to their sensitive cardiac health information hidden in their own Electrocardiograms. Our access control model is able to protect the privacy of users, the confidentiality of medical data, reduce storage space and make it more efficient to upload and download large amount of data.

1. INTRODUCTION

Today, medical data of an individual is often digitized and stored as an electronic health record (EHR) which can be accessed and managed more conveniently, improving efficiency and quality of care [1]. EHRs can also be stored on the cloud, making them available to other parties outside the heath care domain such as government agents, researchers etc. However, designing a secure access control model to protect EHR data is a challenging task. The owner of an EHR should be the main person responsible for regulating access to their health data by creating sets of access control policies to disclose minimum amount of data necessary to different user groups [2]. When EHRs are stored on the cloud, these access policies must also be bound to the data so that access control can be enforced by a distributed mechanism rather than by a centralized access control server [3].

Various cryptography-based solutions [2, 3] have been proposed to control access and protect the confidentiality of EHRs stored on the cloud. However, there are a few limitations inherent in using cryptography. For example, many symmetric key and public key techniques require data owners to manually receive access requests and generate appropriate decryption keys, making it difficult to meet access demands of many interested parties when data owners are not available [2]. Some attribute-based encryption schemes have large ciphertext size and high computation costs depending on the number of attributes used in access policies [3]. Furthermore, resource-intensive encryption techniques might not be favorable in applications such as body sensor networks in which physiological data of patients needs to be continuously monitored and sent securely over the Internet to hospital servers [4]. Therefore, it is desirable to find other access control solutions that require less computing resource and smaller storage space while still maintain a high security and confidentiality level.

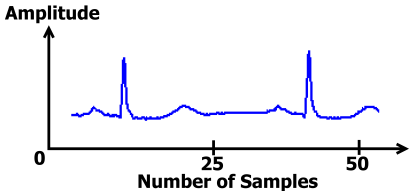

Fig. 1 Normal ECG Waveform

Using steganography is another way to protect the confidentiality of sensitive data. Steganographic techniques allow secret data to be embedded inside another host data such as an image, a text file with insignificant changes to the host data. Only authorized parties possessing correct keys are able to extract the secret data, while unauthorized parties are not even aware of the use of steganography in the host data. Although many steganographic techniques have been proposed in recent years, only a few of those works have addressed issues related to protecting medical data. Kawaguchi et al. [5] created a steganographic access control system to protect digital contents such as images and music files. Nambakhsh et al. [6] invented a novel way to embed ECG signal inside medical images such as X-ray Computed Tomography (CT Scan) and Magnetic Resonance Imaging (MRI). Electrocardiogram (ECG) has also been used as host data. ECG (Fig. 1) measures the electrical activity of the heart over a period of time

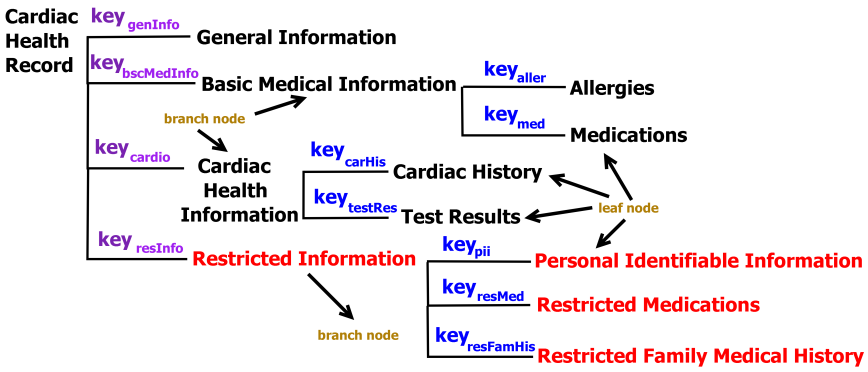

Fig. 2 Medical data hidden in the ECG is organized into a hierarchical tree structure

and is used by cardiologists to diagnose heart diseases. Ibaida et al. [4] proposed a technique to hide patient data inside the ECG. The authors proved that their technique has strong security, low computational complexity and is suitable for resource-constrained devices.

According to our knowledge, there is no research works that apply steganography to create an access control system to protect medical data. Therefore, in this paper, we present a preliminary work on how steganography can be used as an underlying mechanism of an access control system that is applicable in medical context. We focus on an application scenario in which medical information related to the cardiac health of a patient is hidden inside the ECG of that patient. Using steganography, our system is able to give data owners complete control over how their hidden data is retrieved irrespective of wherever the host ECG data is stored.

2. METHODOLOGY

In this section, we describe the design of the steganography-based model to control access to a data structure hidden in the ECG. The steganographic technique is described in Section 2.1. The access control model and the data structure is described in Section 2.2.

2.1. Steganography Technique

At this stage, we select a steganographic technique to hide and retrieve secret information from a host. The technique proposed by Ibaida et al. [4] was chosen because it is directly applicable to ECG. The authors also showed that their technique has strong security, low computational complexity and causes minimal distortion to the host signal.

This technique allows confidential information to be embedded into specific locations called special range numbers of the digital ECG. Some preprocessing parameters are required to transform ECG samples to a set M of non-negative integers The host ECG is then shifted according to the following equation:

$$S = R_{\min} + (M \bmod n) \tag{1}$$

Where S is the set of new shifted values, $R_{\min}$ is the starting value of the target special range, and n is the length of the special range. In our experiments, we used $R_{\min} = 127$, $R_{\max} = 129$ and $n = 3$. The set S of shifted values will be used as a host to hide the secret bit B according to the following equation:

$$M_n = \begin{cases} M_o + (R_{\max} - S) & \text{if } B = 1 \\ M_o - (S - R_{\min}) & \text{if } B = 0 \end{cases} \tag{2}$$

Where M_n is the new resultant value of the host data, M_o is the original host signal sample. Two parameters are needed to extract the hidden data from the host signal: the used range $R_{\min}$, $R_{\max}$ and signal preprocessing parameters. These parameters are used to perform the transformation of the host signal to a set of non-negative integers and then bitwise operation is performed to retrieved the secret bits. There are 2.1475×10^9 available special ranges and 2^{32} possible shifting values. Therefore, without the key, an intruder needs to try $2.1475 \times 10^9 \times 2^{32}$ combinations of special ranges and shifting values to extract the secret message from the host ECG signal.

2.2. Access Control Model

In this section, we describe our steganography-based access control model, including the data structure hidden in the ECG host as well as how access to restricted data is controlled by the data owner using a set of steganographic keys. Our model requires a patient to organize their sensitive cardiac health data into a tree structure as shown in Fig. 2, where each node corresponds to a category of data. After the steganographic process completes, the content of each node can only be accessed by an authorized party possessing the correct

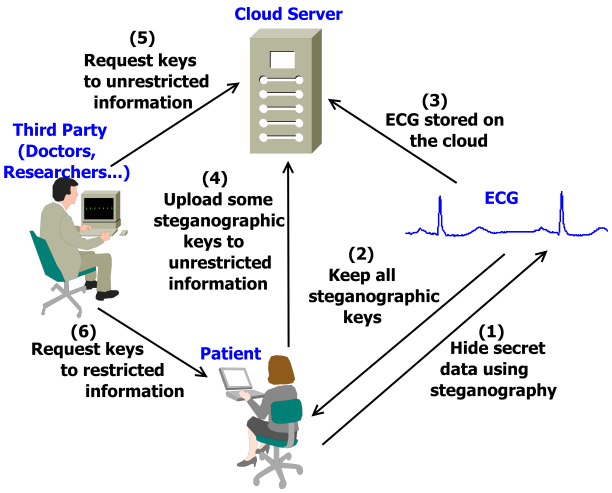

Fig. 3 Using a set of steganographic keys, a patient can control access to information hidden in ECG stored on the cloud

steganographic key to that node. For example, in Fig. 2, the key_{resMed} is required to access the Restricted Medications category. There are two types of nodes, branch node and leaf node. A leaf node has no child node, but a branch node can have an arbitrary number of child nodes. A key to access a branch node can be used to access all child nodes of that node. However, access keys at lower levels cannot be used to access content at higher levels. For example, Fig. 2 shows that key_{cardio} can be used to access content in the Cardiac Health Information node, Cardiac History and Test Results, but key_{carHis} cannot be used to access content in Cardiac Health Information section. This feature allows data owners to have more flexibility in selecting what portions of the hidden data can be accessed by other data users.

Our steganography-based access control model also allows data owners to have complete control over how the secret data hidden in their ECG is accessed, irrespective of wherever the host ECG data is stored. A typical application scenario is shown in Fig. 3. All the steps involved are described as follows:

1. The data owner organizes their data into a hierarchical tree structure and uses steganography to hide the data in their ECG signal.

2. After the steganography process finishes, the data owner will have a set of keys that can be used to retrieve any portion of the hidden data.

3. The host ECG signal can be used for diagnostic purposes or stored in any public cloud-based ECG databases without revealing any information related to the owner.

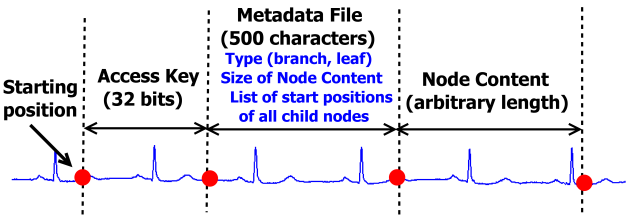

Fig. 4 Structure of a data node hidden in ECG using steganography

4. The data owner can choose to upload some access keys to the cloud if they do not want to restrict access to the sections corresponding to those keys.

5. Interested parties can request some access keys from cloud servers to retrieve some information hidden in the ECG without asking for permissions from the data owner directly.

6. To gain access to restricted information hidden in an ECG signal, keys must be obtained directly from the data owner.

By uploading some keys to unrestricted content hidden in the ECGs, data owners do not need to be available online frequently to approve access requests. All access keys are distributed from the cloud servers or data owners to users via Secure Sockets Layer (SSL).

3. EXPERIMENTS

In our experiments, several ECG hosts were randomly selected from databases that are publicly available on the Internet. For each ECG host signal, experiments were performed to embed and extract data according to the access control model described in Section 2.2. The secret data was a set of text files that were organized into a hierarchical tree as shown in Fig. 2. We assumed that the the ECG host signal size is large enough to store all secret information. All source codes were written in MATLAB programming language and can be grouped into two main modules: embedding module and extraction module.

The embedding module is used by data owners to hide their secret data inside their own ECGs. The steganographic process begins after a random starting position in the ECG signal is automatically selected to hide the first data node. Each subsequent node is located immediately after the first one. Fig. 4 shows how content of a node and its related information are hidden in the ECG host signal. The access key is checked against the one given to a party requesting access. The metadata file contains necessary information to retrieve the node content such as the node type, the size of node content,

and the list of starting positions of all child nodes of that node. The steganographic process finishes after all keys are returned to the users together with the starting position of the metadata node. This node is a special node that contains the list of all other nodes and their starting positions necessary for data extraction. We selected 32 bits to store the key and 500 characters as the maximum length of the metadata file.

Authorized data users need to obtain access keys from data owners or from the cloud to extract the hidden data using the extraction module. These keys are given in a key file which always stores the key and starting position of the metadata node so that the list containing the positions of all other nodes can be extracted. This list is used to locate and extract the keys of nodes hidden in the ECG signal. The extraction module check each key in the key file against the extracted access key of each node. If the key is valid, the content of the node and all its child nodes will be retrieved.

4. DISCUSSIONS

Our steganography-based access control model allows related medical data to be grouped together by embedding them in a host data without significantly affecting the host quality or increasing the size of the host data. Although we only experimented with cardiac health information embedded to ECGs, the model we propose here can also be used in other scenarios such as access control to mental health information embedded in Electroencephalogram (EEG), or information related to a lung cancer embedded in a chest X-ray, for example.

Unlike any cryptographic algorithms, our scheme does not produce ciphertext, thereby reducing storage space and making it more efficient to download and upload a large amount of data. Only authorized parties possessing appropriate access keys are able to retrieve hidden information. Using cryptography, unauthorized parties will not be aware of the existence of data hidden in the host signal, enhancing the confidentiality of sensitive medical data. Our scheme also allows data owners to have complete control over how their hidden data is accessed. Owners can choose to upload to the cloud the host signal and keys to information they do not want to restrict access in order to reduce the number of access requests they need to handle directly. This feature will also allow the sharing of medical data while preserving the owner's privacy and data utility. When data users require more sensitive health information, they need to ask for keys directly from the data owners.

In this paper, we only focus on designing and building a steganography-based access control model to hide and retrieve hidden hierarchical data using a set of access keys. There are other crucial aspects of our scheme that is outside the scope of this paper such as how to delete or modify hidden content using steganography without affecting the quality of the host signal. Key distribution and revocation, access to restricted medical data in emergency cases to save lives are also interesting issues that need to be addressed.

5. CONCLUSIONS

This research has demonstrated how steganography can be used as the main mechanism to design and build an access control model that is applicable in many scenarios to protect the confidentiality of medical data, reduce storage space and make it more efficient to upload and download large amount of data. In future research, we plan to address other technical aspects such as key management and distribution and how to combine our scheme with various cryptography techniques to enhance the security of medical data and privacy of owners.

References

[1] W. Millard, "Electronic health records: promises and realities: a 3-part series part i: the digital sea change, ready or not." *Annals of emergency medicine*, vol. 56, no. 2, p. A17, 2010.

[2] J. Benaloh, M. Chase, E. Horvitz, and K. Lauter, "Patient controlled encryption: ensuring privacy of electronic medical records," in *Proceedings of the 2009 ACM workshop on Cloud computing security*. ACM, 2009, pp. 103–114.

[3] M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption," 2012.

[4] A. Ibaida, I. Khalil, and D. Al-Shammary, "Embedding patients confidential data in ecg signal for healthcare information systems," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 3891–3894.

[5] E. Kawaguchi, M. Maeta, H. Noda, and K. Nozaki, "A model of digital contents access control system using steganographic information hiding scheme," *Frontiers in Artificial Intelligence and Applications*, vol. 154, p. 50, 2007.

[6] M. Nambakhsh, A. Ahmadian, M. Ghavami, R. Dilmaghani, and S. Karimi-Fard, "A novel blind watermarking of ecg signals on medical images using ezw algorithm," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006, pp. 3274–3277.