Towards Identity Management in Healthcare Systems

Silvino Neto¹, Felipe Silva Ferraz¹,², Carlos André Guimarães Ferraz²

¹CESAR
Recife Center for Advanced Studies and Systems
Recife, Brazil
silvino.neto@gmail.com, fsf@cesar.org.br

²Informatics Center
Federal University of Pernambuco
Recife, Brazil
{fsf3, cagf}@cin.ufpe.br

Abstract - Information Systems are of key importance for efficient healthcare services. They improve patient care and administration, providing valuable support for medical diagnosis. To provide such services, healthcare information systems collect and store an extensive volume of patient data in digital format, referred as electronic health record. These records hold a significant amount of patient personal information that may be targeted by cybercriminals. Recently, appalling statistics concerning the exposure and theft of electronic health records have been reported. In this paper we examine the issues related to information privacy and security for healthcare systems and present a new approach for protecting patient data, using an Identity Management framework to preserve patient anonymity. To evaluate this approach, a case study experiment using a disease surveillance platform has been conducted, and its results are exposed in the remainder of this paper.

Keywords: Security; Architecture; Healthcare Systems; Identification; Interoperability; Identity Management.

1 Introduction

Cyber-attacks are responsible for millions of data records stolen every year. Recently, cases of information theft have increased exponentially, in part due to an unprecedented volume of electronic personal information stored, processed or transmitted.

A recent report claims that 1-in-3 electronic medical records have been compromised in 2015 [1]. Health care providers and software vendors admit that almost a hundred million patient records have been exposed last year.

For a cyber criminal, health care data are usually up to 5, 10 or even 50 times more valuable than other forms of personal information. Health care records often include social security and credit card numbers, beyond exploiting the data they contain, criminals may use these records to fill prescription medications and file fraudulent claims [2]. In spite of all the investments made to protect sensitive information, current technologies and solutions still lag behind the discovery of new vulnerabilities and threats by cybercriminals.

In one of our previous work, a disease surveillance middleware platform has been developed, in order to extract statistical information from electronic medical records for monitoring and prediction of outbreak occurrences [3]. Health records are transmitted using the Fast Healthcare Interoperability Resources (FHIR) protocol, which is an international standard for healthcare system-level data exchange. This middleware platform is designed to process thousands of patient records in a highly transactional environment. Patient records usually include both clinical and administrative data. During the course of this research project, we observed several weaknesses related to patient information privacy and identity exposure.

Silva Ferraz and collaborators proposed an experimental security architecture for smart cities, addressing the security flaws aforementioned. This architecture is composed by a software security layer that abstracts communication and cryptography, providing entities (individuals, services or systems) with a mechanism to interact with other systems using unique generated IDs for each corresponding system [4][5]. A significant contribution of this approach is to enable information consumers to manage their own identity across heterogeneous systems [5].

This strategy allows healthcare applications and platforms, which handle sensitive patient information, to prevent patient identity from being exposed, by using a security layer that overrides patient identification with a newly generated correlation ID. Therefore, the patient real identification is preserved, assuring the privacy and anonymity of individuals.

This paper analyzes security issues that may compromise sensitive information in the context of healthcare applications, discusses the strategies to mitigate the risks involved with data leakage, and presents a case study implemented as result of the development of the proposed security architecture, within a disease surveillance middleware platform. The remainder of this paper is organized as follows: Section II sets the scene by discussing the background topics and related research on healthcare systems security. Section III presents a disease surveillance middleware platform as a case study and describes the scenarios evaluated in Section IV. Section V presents our conclusions and future research.

2 Background

Information security and privacy in the healthcare industry is an issue of the utmost importance. The integrity and confidentiality of the information contained in a patient
medical record must be ensured under all circumstances. In this section, we examine security-related aspects for healthcare information systems, addressing topics such as: privacy, systems vulnerabilities and threats.

2.1 Healthcare Privacy and Security

Appari, A. and Eric Johnson, M. noted that privacy is considered as a fundamental principle of the patient-physician relationship [6]. That same principle must be assured by healthcare information systems, in order to guarantee the confidentiality of the data contained in electronic health records. Eventually, a patient’s medical record gathers a significant amount of personal information that may be used for malicious purposes, if stolen or violated by cybercriminals.

Given the emergence of many Internet-based healthcare information systems, the risk of exposure of the information contained in electronic medical records is considerably higher than the one observed for local network healthcare systems. Remote healthcare platforms are likely to accelerate patient care, public health incident reporting, and administrative procedures. However, the challenges presented by remote distribution are much greater in terms of guarantee of data integrity and confidentiality.

The evolution of the healthcare industry has been driven by developments in information technology and regulatory measures. Many countries have introduced legislation in order to assure the privacy of individuals, protecting against the disclosure of personal information [8]. Regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the U.S. and the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada, establish a set of rules to be followed by physicians, hospitals and other health care providers [8][17][18]. HIPAA/PIPEDA aims to ensure that all electronic medical records satisfy certain standards and requirements, regarding documentation, information handling and privacy [7].

Nowadays, electronic health records are constantly exchanged between several parties, being used for a range of purposes [9][12]. In spite of that, there are important questions left to be answered, concerning the ownership and the access privileges over the data collected [7]. As noted by Marci Meingast, Tanya Roosta, and Shankar Sastry, electronic patient records are usually accessible by physicians, insurance providers, and the patients themselves. But it is not clear who has the authority to maintain the information. HIPAA requires that all patients must be granted access to their own electronic medical records, allowing them to correct errors or omissions, and to monitor how their personal information is shared and used [7].

In light of this issue, this paper proposes the use of an Identity Management (IdM) mechanism, allowing patients to maintain and protect their identity, using a security layer that overrides sensitive information when exchanging electronic medical records [8]. This approach is meant to safeguard patient information against security issues like data breaches, identity theft, and unauthorized access. For instance, an insurance provider shall not be capable of visualizing information contained in a patient medical record, other than the reimbursement of medical expenses.

Additional alternatives may be used to improve information privacy and security. Anonymized electronic health records are suitable for data mining, disease surveillance and outbreak monitoring. Access control policies, including partial viewing of specific segments contained in a medical record, can be enforced by the use of a Role-Based Access Control (RBAC) framework [11].

2.2 Healthcare Vulnerabilities and Threats

There are several sources of threats to healthcare information privacy and security. Internet-based healthcare information systems are more likely to be exposed by cybercriminals compared to regular paper-based medical records. Sensor data, which consist of data captured and transmitted by wearable medical devices for remote patient monitoring, are also susceptible to eavesdropping and skimming attacks, when it’s transmitted wirelessly. Data access, storage, and integrity are key challenges when implementing electronic health records and in-home sensor networks [7]. These are some examples of the threats that endanger the privacy and security of the patient information exchanged by healthcare systems. Past studies have broadly separated privacy threats and vulnerabilities into two main categories [6], presented as follows:

1) Organizational Threats: These are situations that occur as result of inappropriate access of patient information by either internal or external agents [6]. For example, a hospital employee that abuse their access privileges to view patient data for no justifiable reason, or an outside attacker that infiltrates a hospital healthcare information systems in order to steal a patient medical record. Both cases are characterized as organizational threats, given that patient information has been violated due to the exploit of vulnerabilities in the organization itself. These attacks may be inflicted by an individual with modest financial backing, or very elaborated well-funded groups [6]. Studies suggest that organizational threats may be further categorized into five different levels, presented in increasing order of sophistication [6]:

Accidental disclosure: Healthcare personnel unintentionally disclose patient information to others.
Insider curiosity: An insider with data-access privilege pries upon a patient’s records out of curiosity or for their own purpose.
Data breach by insider: Insiders access patient information and transmit it to outsiders for profit or revenge.
Data breach by outsider with physical intrusion: An outsider enters the physical facility either by coercion or forced entry and gains access to the system.
Unauthorized intrusion of network system: An outsider, including former employees, patients, or hackers, intrudes upon a patient's records out of curiosity or for their own purpose.
into an organization’s network from the outside to gain access to patient information or render the system inoperable.

2) Systemic Threats: These are risks related to the exploit of the information flow chain, perpetrated by legally authorized personnel that use the disclosed information beyond its original intent [6]. For instance, an employer that denies a promotion based on the information contained in the employee’s medical records, or a health insurance company that denies coverage due to a preexisting medical condition.

3 Case Study

Previous studies brought into attention the need to make further improvements related to information security on healthcare information systems [5] [6] [7] [8]. Based on this need, this section describes the use of a Healthcare Security Layer (HSL) developed on top of the Platform for Real-Time Verification of Epidemic Notification (PREVENT), which is a disease surveillance middleware platform used to process statistics extracted from electronic medical records in order to anticipate and report outbreak occurrences [3].

3.1 Middleware Architecture

PREVENT is a Message-oriented Middleware (MOM) platform, built to collect and process a large volume of information, in a highly scalable fashion. It provides a set of RESTful interfaces to be used by healthcare information systems to exchange aggregate data reports held in electronic medical records [3].

PREVENT handles a significant amount of sensitive patient information. However, current support for data security mechanisms is limited to the use of the TLS/SSL data transport protocol [15][16]. Therefore, numerous security breaches have been observed in this platform, most of them related to privacy concerns. In order to address this situation, this case study examines the use of a Healthcare Security Layer to be integrated into PREVENT architecture.

The HSL aims to assure patient anonymity, protecting the privacy and confidentiality of individuals, by overriding sensitive information included in electronic medical records with new identifiers, generated from the combination of the patient personally identifiable information (e.g. Name, Social Security, and Credit Card Numbers, etc.) and the healthcare provider identification.

The diagram exhibited in Figure 1 is an overview of PREVENT multilayered architecture, including its most significant software components. PREVENT is composed by 5 different layers, each one of them with a distinct set of responsibilities, described as follows:

**Presentation Layer**: Hosts the RESTful Servlet endpoints used to receive messages and subscription requests dispatched by healthcare information systems.

**Service Layer**: Extracts and processes the data contained in the electronic medical records received from healthcare information systems, and coordinates communication with other layers in order to complete data analysis operations.

**Persistence Layer**: Stores and retrieves information persisted in a NoSQL database.

**Communication Layer**: Handles communication between PREVENT and the Google Cloud Publish/Subscribe Services API.

**Security Layer**: Encapsulates the access to PREVENT Communication Layer and provides a mechanism to ensure the privacy and confidentiality of patient data, using an Identity Manager component.

The physical architecture of this platform is presented in Figure 2. In this deployment diagram, it is possible to observe that communication between PREVENT and its subscribed healthcare information systems is performed using HTTP over TLS/SSL [15][16].

In spite of the fact that the information exchanged between PREVENT and other healthcare systems is transmitted over a secure channel, when the data have reached the other end, they may be exposed and violated by a third party. In other terms, once the data is out of the scope of this platform, privacy and security concerns cannot be enforced, compromising the anonymity of patients. In order to address this issue, in the next section we discuss the use...

3.2 Healthcare Security Layer

Silva Ferraz et al. observed that, in the context of Smart Cities, citizen identity integration across multiple system platforms and services comprises the goal of allowing the citizen to manage their own identity. This also includes what type of information is released about them to whom or when, while anonymous, aggregate data are more widely available [5]. That same concept can be translated and applied here, in relation to the identity of a patient.

The HSL is responsible for overriding patient sensitive information contained in messages received from healthcare providers, prior to their delivery to subscribed healthcare information systems. Message information replacement occurs as part of a cryptographic procedure, enabling the protection of patient personally identifiable information. The new identifiers are generated by a combination of patient identifiers and healthcare provider ID, being unique within the whole platform [5].

In order to generate a new patient identifier, PREVENT concatenates the original identifier value with the healthcare provider ID, the resulting string value is then ciphered using a 256-bits cryptographic hash function (SHA-2) [20]. Finally, the obtained hash byte array is encoded back to string format using Base64 binary-to-text representation [19]. The new identifiers are then used to replace the original values in the messages received in FHIR/JSON format.

Also, a new entry is persisted in the platform NoSQL database, in order to correlate the original patient identifier, the healthcare provider ID, and the newly generated identifier. In order to illustrate the HSL data flow, a collaboration diagram is exhibited in Figure 3. This diagram presents a scenario where Hospital A wants to share electronic medical records with PREVENT middleware platform for aggregate data reporting. The patient records received are processed and delegated to PREVENT Communication Layer, which checks whether the received patient identifiers and the healthcare provider ID are related to previously generated hashes. According to the outcome of this test, hash IDs may be created or retrieved, in order to override patient information. At last, the updated messages are dispatched to Hospital B.

3.3 Electronic Health Records

As previously discussed in this paper, electronic health records are composed by a set of sensitive personal information. Recently, the numbers of incidents related to patient data leakage and theft have increased significantly. This situation exposes the need for more robust and sophisticated information security. PREVENT’s HSL aims to improve information privacy by using two combined techniques: cryptography and anonymity. In this section, we demonstrate how it affects the contents of an electronic healthcare record, and what type of data should typically be anonymized.

A regular electronic patient record typically contain a significant amount of personal information, such as: full name, home address, phone number, social security number, and sometimes even a credit card number may be present. If data is exploited for malicious purposes, it may lead to a number of damages and consequences for the patient. Using a secure channel for message transmission ensures that the data within a patient record will not be compromised. However, when the data is delivered to another healthcare information system, it may be an easy target for violation or theft, as previously demonstrated in this paper. In the following sections, this paper presents an assessment of this platform, introducing a comparison between a secure and non-secure approach for healthcare data exchange, in terms of functional and performance-related results.

3.4 Non-Secure Approach for Data Exchange

In this approach, electronic health records are exchanged between this platform and the subscribed healthcare information systems, without any type of modification. Therefore, any personal information held in the patient record, is sent along with aggregate clinical data reports. Towards establishing a reference parameter for functional and performance-related results, in this section we analyze PREVENT functional behavior, and collect a few metrics in order to assess the performance of this middleware platform. The results gathered will set the boundaries for comparison between the secure and non-secure approaches.

This assessment is performed at the Google Cloud Platform, which is a modular cloud hosting service. The machine type used for this test is a standard Google Virtual Machine Instance, with a single virtual CPU that is equivalent to a 2.6GHz Intel Xeon E5 and 3.75GB of memory. The PREVENT middleware application is deployed at the Google App Engine, which is an application server for hosting cloud-based web applications [13][14]. The test cases used for this assessment are distributed amongst four different testing scenarios. The first and the second scenarios are composed by 300 message samples.
The third and the fourth scenarios are comprised by 500 messages each. A total of 1600 messages are expected to be sent to the middleware application using the Apache JMeter, which is a Java-based performance testing tool. Each message holds information related to a unique patient. The platform is configured to deliver messages to 50 registered HTTPS endpoints, each one of them acting as healthcare information systems. Each registered HTTPS endpoint is either a Java Servlet class deployed at the Google App Engine or a PHP file hosted at the Digital Ocean NGINX server that simply logs the current timestamps and the contents of each FHIR message received.

FHIR health records are usually represented in JSON format. They are composed by several key-value pairs that are used for data representation. According to the data structure of a FHIR message, sensitive patient information such as: home address, identifier value (social security number), and full name are stored in plain text in PREVENT NoSQL database. Later, the FHIR messages persisted are also delivered to the platform subscribers without any type of type of treatment for data anonymization. In the course of this assessment, this middleware platform presented a stable functional behavior. No messages were reported lost nor have any requests been rejected. All messages have been successfully delivered, in accordance with the expected outputs for this simulation. Table 1 illustrates a set of metrics that has been collected in order to analyze performance-related results. The results gathered during this assessment are an essential input, allowing us to identify potential performance bottlenecks related to the use of cryptographic algorithms for data anonymization.

### TABLE I. PERFORMANCE METRICS FOR NON-SECURE APPROACH

<table>
<thead>
<tr>
<th>Number of Samples</th>
<th>Median (ms)</th>
<th>Throughput</th>
<th>Lost Messages</th>
<th>Failed Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>231</td>
<td>211.942/Min</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>218</td>
<td>218.125/Min</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>209</td>
<td>257.889/Min</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

#### 3.5 Secure Approach for Data Exchange

This approach relies on the use of cryptography and encryption, in order to anonymize electronic health records exchanged between this platform and the registered healthcare information systems. Hence, all personal information included in the patient medical records must be ciphered and encoded prior to its storage and delivery. In the previous section, we have established reference parameters for functional and non-functional requirements, based on the results collected during the assessment of the non-secure approach. In the scope of this assessment, the same set of metrics is to be collected, for later comparison against the metrics previously obtained. The differences observed over the collected results are an accurate indicator of how significant is the impact for performance and functionality. The present assessment is to be conducted using precisely the same testing environment, as the one utilized for the non-secure approach. Furthermore, an updated version of the PREVENT middleware application has been deployed at the Google App Engine, including the security features provided by the Healthcare Security Layer [13][14]. The testing script to be used for this assessment is the same as the one presented for the previous experiment, including testing tools, methods, and datasets. Four rounds of execution are programmed for this evaluation, and the same amount of message samples will be used for processing.

As previously discussed in this paper, FHIR health records are received by this platform over a secure channel. However, sensitive patient information held on these records is usually received in plain text format. That is, as long as the received information is kept within the boundaries of this disease surveillance platform, patient information privacy and security is preserved. The HSL aims to address this limitation by using an Identity Management framework that overrides sensitive patient information with ciphered values for outgoing messages [10]. The HSL allows system administrators to define a set of attributes contained in a patient medical record that are set to be encrypted for protection. Also, a new NoSQL entity kind named GeneratedHashValues was created in order to store the generated hash IDs, indexed by the combination of the patient attribute value and the healthcare provider ID. To demonstrate the HSL capabilities for this assessment, we have enabled the protection of two health records attributes: patient identifier (social security number) and full name. It is important to observe that this assessment will likely reproduce the worst-case scenario, given that an empty dataset for generated hash IDs will be used. Put simply, for every message received, both identifiers (social security number and full name) will be encrypted and stored, resulting in additional processing overhead.

![Figure 4](image.png)

Figure 4 shows a FHIR health record snippet, obtained prior to data anonymization and protection. Several attributes were suppressed for the sake of simplicity and due to the size limitations of this paper. As previously discussed, under the non-secure approach, patient data is represented in human-readable format. After the encryption of sensitive patient information, a secure and privacy-assured patient...
record is obtained, as the one exhibited in Figure 5. As occurred for the non-secure approach, the middleware sustained a stable functional behavior during the whole experiment. No lost messages or functional errors have been observed. As presented in Table 2, the same set of metrics has been collected for further comparison.

![Figure 5. FHIR Health Record Snippet After Anonymization.](image)

In the next section, we compare and analyze the results obtained by both approaches, in order to establish the balance between the following QoS requirements: performance, privacy, scalability, and security.

**TABLE II. PERFORMANCE METRICS FOR SECURE APPROACH**

<table>
<thead>
<tr>
<th>Number of Samples</th>
<th>Median (ms)</th>
<th>Throughput</th>
<th>Lost Messages</th>
<th>Failed Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>230</td>
<td>205.590/min</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>300</td>
<td>232</td>
<td>235.331/min</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>229</td>
<td>248.063/min</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>233</td>
<td>238.226/min</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**TABLE III. PERFORMANCE METRICS COMPARISON**

<table>
<thead>
<tr>
<th></th>
<th>Non-Secure Approach</th>
<th>Secure Approach</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Throughput</td>
<td>231 ms 211.942/min</td>
<td>230 ms 205.590/min</td>
<td>Diff. (%)</td>
</tr>
<tr>
<td></td>
<td>218 ms 218.125/min</td>
<td>232 ms 235.331/min</td>
<td>-6.42 -1.07</td>
</tr>
<tr>
<td></td>
<td>209 ms 257.889/min</td>
<td>229 ms 248.063/min</td>
<td>9.56 3.81</td>
</tr>
<tr>
<td></td>
<td>211 ms 250.077/min</td>
<td>233 ms 238.226/min</td>
<td>10.4 4.73</td>
</tr>
</tbody>
</table>

These experiments all used the PREVENT middleware platform for processing electronic health records received for outbreak detection and anticipation. PREVENT was running over the Google Cloud Platform, with fifty randomly chosen subscribers for message delivery simulation. Throughout both experiments, secure and non-secure approaches did not differ at all, in terms of functional behavior. All messages were successfully delivered after the completion of both experiments. This is a compelling evidence of the HSL transparent functional operation. Similarly, minor performance differences have been observed, when comparing the results of the secure and non-secure strategies. Tables I and II present the results obtained for each approach, in an isolated manner. Table III presents a comparison between the results obtained for both approaches, in a combined manner. In spite of the fact that our secure approach performs an additional cryptographic operation for patient information. The performance impact observed is a very insignificant decline, especially when measuring the benefits of using a security-based approach for patient data. After analyzing the measured numbers, we can conclude that the performance levels of our secure approach for healthcare data exchange was similar than the one observed for the common insecure strategy, given that in general the throughput and the median response time were slightly impacted. As security positive impacts we can point out the following:

**Privacy increased:** Now patient data are no longer associated with the patient real identity. Therefore, it is not possible to link a certain disease or condition to one specific individual.

**Anonymity:** In accordance with the previous bullet statement, the separation between patient identity and its data promotes a transparent and immediate mechanism for data anonymization, by simply removing information from the security layer.

**Patient tracking:** The papers [21] and [22] states that citizen tracking can be harmful, given the risks associated with sensitive data exposure. Applying that same principle to patient tracking, through the adoption of the proposed solution, patient tracking is no longer viable.

**Authentication and Authorization enforcement:** The presented solution is mainly focused on Identity Management, nevertheless it provides the means to re-enforce eventual authentication and authorization needs through its separated approach. Hence, in case a user or physician, for instance, has no appropriate permission to do so, they will not be granted access to the patient information.

It should be remembered that, this assessment was performed in simulation of the worst-case scenario for the secure approach, given that we used an empty dataset for generated hash IDs. Therefore all protected identifiers, contained in every message received, had to be encrypted and stored prior to message delivery.

**5 Conclusions**

This paper has presented a case study focused on the adoption of a security architecture designed for the improvement of information privacy and security. The proposed architecture has been originally envisioned as a
response to privacy and security concerns observed in the context of Smart Cities. The security challenges encountered for smart-city environments are similar, to some extent, to the ones observed for healthcare information systems. In order to address this situation, the present research performed a case study based on the development of a Healthcare Security Layer (HSL) within an existing disease surveillance middleware platform, used to extract statistical information from electronic medical records for monitoring and prediction of outbreak occurrences. The HSL is comprised of an Identity Management framework to some extent, to the ones observed for healthcare information systems. In order to address this situation, the present research performed a case study based on the development of a Healthcare Security Layer (HSL) within an existing disease surveillance middleware platform, used to extract statistical information from electronic medical records for monitoring and prediction of outbreak occurrences. The HSL is comprised of an Identity Management framework and how to use it in a real-time disease surveillance system.

The results gathered in the course of this case study demonstrate the substantial benefits of using an Identity Management platform for enforcing privacy and security requirements for healthcare systems. It should be highlighted that transparent functional behavior has been observed during the whole experiment, and only a minor performance impact has been noted. As future work it is planned to conduct a more detailed performance and load validation in order to assess the impact of an extra layer in a performance based environment, also it is planned to connect the solution with different data bases and services, that works as data sources, validating the interoperable characteristic of the security layer.

6 References


