Chapter 38
Medical Data Analytics in the Cloud Using Homomorphic Encryption

Övünç Kocabaş
University of Rochester, USA

Tolga Soyata
University of Rochester, USA

ABSTRACT

Transitioning US healthcare into the digital era is necessary to reduce operational costs at Healthcare Organizations (HCO) and provide better diagnostic tools for healthcare professionals by making digital patient data available in a timely fashion. Such a transition requires that the Personal Health Information (PHI) is protected in three different phases of the manipulation of digital patient data: 1) Acquisition, 2) Storage, and 3) Computation. While being able to perform analytics or using such PHI for long-term health monitoring can have significant positive impacts on the quality of healthcare, securing PHI in each one of these phases presents unique challenges in each phase. While established encryption techniques, such as Advanced Encryption Standard (AES), can secure PHI in Phases 1 (acquisition) and 2 (storage), they can only assure secure storage. Assuring the data privacy in Phase 3 (computation) is much more challenging, since there exists no method to perform computations, such as analytics and long-term health monitoring, on encrypted data efficiently. In this chapter, the authors study one emerging encryption technique, called Fully Homomorphic Encryption (FHE), as a candidate to perform secure analytics and monitoring on PHI in Phase 3. While FHE is in its developing stages and a mainstream application of it to general healthcare applications may take years to be established, the authors conduct a feasibility study of its application to long-term patient monitoring via cloud-based ECG data acquisition through existing ECG acquisition devices.

DOI: 10.4018/978-1-4666-8756-1.ch038

INTRODUCTION

Utilizing cloud computing resources such as Amazon EC2 (Amazon, n.d.), Microsoft Azure (Microsoft, n.d.), or Google (Google, n.d.) is commonplace for many corporations, due to its ability to prevent vast infrastructure investments. This concept dates back to the beginning of the Internet boom more than a decade ago with the emergence of the *Application Service Provider (ASP)* model: Rather than making an investment in costly server hardware, software licensing fees, and the personnel to manage this infrastructure, corporations can *rent* computation time, storage space, and licensing fees by running such applications as Salesforce.com (Salesforce, n.d.) over the Internet. The ASP model prevents upfront costs: a monthly subscription fee and a flexible licensing scheme allows smaller corporations to immediately start using such programs and expand with virtually no boundaries, since the computational and storage resources are provided by the application service provider (ASP) and the ASP can pool resources for many other clients. Additionally, this eliminates the need for corporations to have any expertise in setting up such sophisticated server infrastructure and the training on the application is done through online seminars.

Another dramatic example of such an ASP model is Paypal (Paypal, n.d.). The introduction of a merchant Application Programming Interface (API) by Paypal allowed any size corporation to start their business with near-zero investment, accept payments over the Internet by using Paypal as the intermediary, and grow with virtually no boundary. These examples show that, it is natural to shift the responsibility of computing (and storage) infrastructure investments to operators that can deliver their services by using the Internet as the delivery channel (i.e., Cloud Operators). By virtualizing their computational and storage resources, these cloud operators can provide these resources to their customers at a fraction of what the customers can build them for.

While endless examples exist for such generic cloud computing offerings, one area that can benefit significantly from it deserves specific attention: Medical cloud computing. When the data storage is outsourced to a cloud operator over the Internet, an important issue arises: data privacy. Although different applications have different sensitivity levels to this issue, the highest level of sensitivity is clearly in the medical arena (Kocabas et al, 2013). Personal Health Information (PHI) is one of the most scrutinized concepts, protected by laws and regulations of the U.S.A. The Health Insurance Portability and Accountability Act (HIPAA, n.d.) dictates a strict set of rules and regulations to prevent the PHI from being misused. Therefore, to expand the cloud computing into the medical arena, one must clearly formulate the entire concept around these restrictions.

Cloud computing is an active research area for medical applications, partly due to the push by the US government to modernize the US Health system (Lobodzinski & Laks, 2012). The motivations behind this move are: 1) improving the quality of healthcare by using additional cloud-based long-term patient monitoring data that are otherwise unavailable to the healthcare professionals, and 2) reducing the operational costs at healthcare organizations (HCO) by eliminating the datacenters operated by HCOs. Long-term patient monitoring data (e.g., patient vitals such as ECG and blood pressure), obtained by sensors that transmit their patient information over the cloud can be used as an auxiliary diagnostic tool to improve diagnostic accuracy. This expands the boundaries of an HCO to outside the HCO by allowing the patients to use long-term monitoring devices, such as ECG patches.

In this chapter, we study the feasibility of such a cloud-based long-term monitoring system while preserving PHI. Preserving PHI requires ensuring data privacy at three distinct phases: Phase I. Acquisition, is where the medical data is acquired from a patient, whether it is within the HCO, or outside the HCO via disposable devices such as