Expert Systems with Applications 42 (2015) 7905–7916

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# CloudID: Trustworthy cloud-based and cross-enterprise biometric identification

Mohammad Haghighat [a,*], Saman Zonouz [b], Mohamed Abdel-Mottaleb [a]

[a] Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33146, USA
[b] Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ 08854, USA

ARTICLE INFO

*Article history:*
Available online 24 June 2015

*Keywords:*
Biometric identification
Cloud security
Encrypted biometrics
Face recognition
Search over encrypted data

ABSTRACT

In biometric identification systems, the biometric database is typically stored in a trusted server, which is also responsible for performing the identification process. However, a standalone server may not be able to provide enough storage and processing power for large databases. Nowadays, cloud computing and storage solutions have provided users and enterprises with various capabilities to store and process their data in third-party data centers. However, maintenance of the confidentiality and integrity of sensitive data requires trustworthy solutions for storage and processing of data with proven zero information leakage. In this paper, we present CloudID, a privacy-preserving cloud-based and cross-enterprise biometric identification solution. It links the confidential information of the users to their biometrics and stores it in an encrypted fashion. Making use of a searchable encryption technique, biometric identification is performed in encrypted domain to make sure that the cloud provider or potential attackers do not gain access to any sensitive data or even the contents of the individual queries. In order to create encrypted search queries, we propose a k-d tree structure in the core of the searchable encryption. This helps not only in handling the biometrics variations in encrypted domain, but also in improving the overall performance of the system. Our proposed approach is the first cloud-based biometric identification system with a proven zero data disclosure possibility. It allows different enterprises to perform biometric identification on a single database without revealing any sensitive information. Our experimental results show that CloudID performs the identification of clients with high accuracy and minimal overhead and proven zero data disclosure.



## 1. Introduction

Substitution of biometrics for passwords in authentication and identification systems received attention in security systems (Jain, Ross, & Pankanti, 2006). Biometric identifiers are distinctive and measurable characteristics used to recognize individuals (Jain, Hong, & Pankanti, 2000a). Some of the well-known biometrics used for human identification are fingerprints, face, iris, voice and DNA. Some of the advantages of biometrics over passwords include their higher level of security, mobility, difficulty to forge, and user friendliness. According to a new study published by Javelin Research (Javelin Strategy & Research, 2014), smartphone users prefer using biometrics as an alternative for passwords, which brings more security to new technologies such as *Apple Pay*. In spite of all these advantages, there are a few challenges that biometric systems face.

One of the major challenges of biometric systems is the variability in the characteristics of the biometrics for each individual. Human face, as an example biometric trait, is a complex object with features that change over time. Facial features change due to changes in illumination, head pose, facial expressions, cosmetics, aging, and occlusions because of beard or glasses. However, we humans have an ability to recognize faces and identify persons at a glance. This natural ability does not exist in machines; therefore, we design intelligent and expert systems that can simulate the recognition artificially (Khashman, 2008). These intelligent systems are trained using the biometric information of the subjects who are enrolled in the system. The identification is performed through a comparison between the biometric information of the query subject and the enrolled subjects. Therefore, these systems need to store biometric information of all enrolled subjects in databases to be utilized at the time of query.

As the number of subjects increases, the system requires more storage capacity and more processing power. On the other hand, these databases need to be accessible by all enterprises that make use of biometric identification. The need to be accessed by multiple

* Corresponding author.
*E-mail addresses:* haghighat@umiami.edu (M. Haghighat), saman.zonouz@rutgers.edu (S. Zonouz), mottaleb@miami.edu (M. Abdel-Mottaleb).

http://dx.doi.org/10.1016/j.eswa.2015.06.025
0957-4174/

enterprises and to have a high processing power motivate the use of a cloud-based system to store and process the data. The work presented in this paper is a key step towards a cloud-based unified storage system for personal records. The idea is to create an encrypted database of personal records of individuals, *e.g.*, name, date of birth, educational information, banking or credit history, medical records, criminal records, insurance, etc., as a unified and privacy-preserving cloud-based database. Biometric information of individuals are used as a key attribute for this database.

The growing popularity of cloud-based systems has increased the importance of addressing the serious issue of the security of the data stored in the cloud (Fernandes, Soares, Gomes, Freire, & Inácio, 2014; Kandukuri, Paturi, & Rakshit, 2009; Padilha & Pedone, 2015; Ren, Wang, & Wang, 2012; Takabi, Joshi, & Ahn, 2010). In case of using biometrics to have access to records stored on the cloud, there is the risk of identity theft, because biometric data of the enrolled subjects can be stolen and misused against their will. The biometric data is unique and irrevocable, and unlike passwords users cannot change their biometrics. Consequently, the system must guarantee the preservation of the users' privacy, and therefore, the biometric database has to be encrypted.

Since the encrypted biometric database is stored in a public cloud, the identification process should be done with minimum amount of information leakage. That is, the comparisons need to be performed without the decryption of the data to prevent eavesdroppers from any access. However, the variations in the biometrics of each subject bring about a serious problem in the encrypted domain. Small changes in the data (plaintext) result in big differences in the ciphertext (encrypted plaintext). This difference can mislead the recognition process. Consequently, it is not feasible to just simply add an encryption scheme to a biometric identification system in order to secure the data and expect to obtain the same results that are obtained without the encryption.

***Contributions:*** In this paper, we present a privacy-preserving cloud-based identification system (CloudID), which allows users to securely store their confidential information in untrusted public clouds. It gives them the opportunity of having effective and secure storage along with the computational power of the cloud infrastructures as well as controlled and flexible data access by multiple agents. Our proposed approach is the first cloud-based biometric identification system with a proven zero data disclosure possibility. CloudID performs the identification process in the encrypted domain without decrypting the data. This prevents the cloud provider or potential attackers from gaining access to any sensitive data or even the contents of the individual queries.

Unlike other privacy-preserving biometric identification methods, our approach does not apply a distance-based matching. However, using the query sample, it creates an encrypted conjunctive range query, which is applied on the encrypted gallery samples stored in the cloud. In this scenario, the only revealed piece of information is the binary matching result, *i.e.*, *match* or *not match*. This makes CloudID secure against *center search attack* (Pagnin, Dimitrakakis, Abidin, & Mitrokotsa, 2014) in which the attacker can recover the biometric template even if it is stored encrypted. In order to improve the performance of the biometric-based identification in the encrypted domain, we propose a k-d tree structure to create encrypted search queries. Applying this structure in the core of a searchable encryption technique helps the system not only to quantize the biometric features but also to handle the variations in the biometric data. Moreover, our algorithm is not limited to any specific type of biometric data and it can work with any biometric trait and any feature extraction method.

A working prototype of the CloudID framework is implemented and evaluated using a public biometric database. Our experimental results show the feasibility of CloudID for accurate biometric identification with no confidential data disclosure possibility, which enables building trustworthy storage systems for sensitive records.

The rest of the paper is organized as follows. Section 2 presents related work from the literature. Section 3 provides a brief overview of the system whose complete design is described in Sections 4 and 5. The implementation details and overall performance of the system are presented in Section 6. Finally, Section 7 concludes the paper.

## 2. Related work

In the literature, there are two sets of studies considering the privacy-preserving biometrics identification. The first set of studies (Barni et al., 2010; Blanton & Gasti, 2011; Erkin et al., 2009; Huang, Malka, Evans, & Katz, 2011; Osadchy, Pinkas, Jarrous, & Moskovich, 2010, 2013; Sadeghi, Schneider, & Wehrenberg, 2009) only achieve privacy-preserving at the time of executing the query, protecting the confidentiality of both server and client. In these approaches, the server, in which the biometric database is stored, is considered to be trusted, and the biometric database is stored unencrypted. This allows the server to have access to the contents of the biometric database. However, these approaches cannot be used in case of untrusted servers such as clouds. In the second set of studies, the server is not trusted and the biometric database is encrypted (Bringer, Chabanne, & Kindarji, 2009, 2011, 2013a, 2013b, 2014). However, these algorithms have some limitations and suffer from information leakage. In this section, we briefly describe these methods and compare them with our proposed approach.

To the best of our knowledge, Erkin et al. (2009) considered the problem of privacy preserving biometric identification for the first time. They proposed a privacy-preserving face recognition system based on the well-known eigenface approach introduced by Turk and Pentland (1991a, 1991b). They employed Pailliers cryptosystem (Paillier, 1999), as an additive homomorphic encryption and calculated the Euclidean distance between face image feature vector from client and server's face image database. The matching algorithm is performed between the client and the server without revealing the client's biometric information or the result of the query to the server. At the same time, the client cannot learn from the database stored in the server. Later, Barni et al. (2010) proposed a similar algorithm for a fingerprint recognition system, *FingerCodes* (Jain, Prabhakar, Hong, & Pankanti, 2000b). Both of these protocols (Barni et al., 2010; Erkin et al., 2009) rely on homomorphic encryption and do not try to find the specific match but the group of the nearest matches within some threshold.

Sadeghi et al. (2009) improved the efficiency of Barni's algorithm by applying a hybrid approach where *garbled circuits* were used in conjunction with homomorphic encryption to find the minimum distance. Huang et al. (2011) combined the algorithms proposed in Sadeghi et al. (2009) and Erkin et al. (2009) to further improve the computational and bandwidth efficiency of the system.

The main idea in Erkin et al. (2009), Sadeghi et al. (2009) and Huang et al. (2011) is to find the nearest match for a query in the biometrics database based upon the Euclidean distance. In these references, each query is encrypted using the public key published by the client and sent to the server. The server also encrypts each biometric data in the database using an additive homomorphic encryption using the same public key. Then, the Euclidean distances between the query and each gallery in the database are calculated in the encrypted domain, $d_1, d_2, \ldots, d_n$. In Sadeghi et al. (2009) and Huang et al. (2011), this information is fed into a garbled circuit, which finds the closest match by calculating $i^* = argmin_i(d_1, d_2, \ldots, d_n)$.