2023-05-23

Medical Systems Data Security and Biometric Authentication in Public Cloud Servers

Santos, N

10.1109/tetc.2023.3271957
IEEE Transactions on Emerging Topics in Computing
Institute of Electrical and Electronics Engineers (IEEE)

Medical Systems Data Security and Biometric Authentication in Public Cloud Servers

Nelson Santos, Bogdan Ghita, and Giovanni Masala

Abstract— Advances in distributed computing and virtualization allowed cloud computing to establish itself as a popular data management and storage option for organizations. However, unclear safeguards, practices, as well as the evolution of legislation around privacy and data protection, contribute to data security being one of the main concerns in adopting this paradigm. Another important aspect hindering the absolute success of cloud computing is the ability to ensure the digital identity of users and protect the virtual environment through logical access controls while avoiding the compromise of its authentication mechanism or storage medium. Therefore, this paper proposes a system that addresses data security wherein unauthorized access to data stored in a public cloud is prevented by applying a fragmentation technique and a NoSQL database. Moreover, a system for managing and authenticating users with multimodal biometrics is also suggested along with a mechanism to ensure the protection of biometric features. When compared with encryption, the proposed fragmentation method indicates better latency performance, highlighting its strong potential use-case in environments with lower latency requirements such as the healthcare IT infrastructure.

Index Terms— Data Fragmentation, Cloud Security, NoSQL Database, Security and Protection

1 Introduction

THE progressive trend of shifting the burden and associated costs of infrastructure procurement, maintenance, personnel, as well as the focus on core tasks related to the system or service being delivered, to Cloud Service Providers (CSPs) while paying for infrastructure according to its usage [1], has increased the popularity of cloud computing with its accelerated adoption across industries and organizations of all sizes.

Currently, existing cloud computing deployment models are categorized based on where the underlying infrastructure resides and who controls it. These models include public cloud which is, as suggested by the name, provisioned to be accessed by all users while the infrastructure is hosted and operated by a CSP. Meanwhile, in private clouds, the infrastructure is provisioned to a single entity and its associated units. Community cloud, on the other hand, is when the infrastructure is provisioned to a community of consumers from entities with similar objectives or concerns. Finally, a hybrid cloud is a combination of two or more distinct deployment models which are interconnected either by standardized operational methodologies or its data and applications which are interoperable between models [2].

Cloud computing can be categorized as: (1) Software as a Service wherein end-users consume applications running on cloud infrastructure through a thin client interface such as a web browser; (2) Platform as a Service where application platforms (middleware, web servers, and databases) are built and maintained by the provider while the organization focuses on the creation, development, and packaging of software bundles; and (3) Infrastructure as a Service where CSPs give the capability of provisioning computing resources and operational control over its operating system, storage, network, and applications to organizations without ceding control over the underlying infrastructure [2].

Furthermore, cloud computing radically reduces entry costs associated with running and hosting compute-intensive tasks. Its adaptive infrastructure is shared across different user groups who are completely segregated, while its resources are balanced according to demand. As a result, the immediate provision of resources lowers barriers for small organizations to expedite the time-to-market of products and offerings. At the same time, enterprise organizations can also easily scale their services and tailor provision based on client demand [3], easing therefore the burden associated with hosting data centers.

Although the cloud computing paradigm has offered many benefits, there are also many barriers that hinder its adoption. Among such concerns is maintaining the security and privacy of data residing in the cloud. This is because by adopting this concept, organizations cede authority and control of data over to CSPs, resulting in various issues that endanger data security. In some cloud models, organizations rely heavily on CSPs to store and safeguard their data, despite the fact that CSPs oftentimes do not publicly disclose their method of handling data and storage procedures [4-6]. Companies are therefore unable to apply specific mechanisms on the data, such as encryption or

- *Nelson Santos is with the School of Engineering, Computing and Mathematics, University of Plymouth, Plymouth, United Kingdom. E-mail: nelson.santos@plymouth.ac.uk.*
- *Bogdan Ghita is with the Centre for Security, Communications and Network Research, University of Plymouth, Plymouth, United Kingdom. E-mail: Bogdan.ghita@plymouth.ac.uk.*
- *Giovanni Masala is with the School of Computing, University of Kent, Canterbury, United Kingdom E-mail: g.masala@kent.ac.uk.*

xxxx-xxxx/0x/$xx.00 © 200x IEEE Published by the IEEE Computer Society

access control. Moreover, the multi-tenancy and resource-sharing design of cloud computing fall short in determining and segregating the reputation of occupants. In other words, individuals and organizations with different interests share the same physical infrastructure, and in the unfortunate event where part of the stored data is attacked, the perpetrator may propagate the attack to impact other resources through guest-hopping or side-channel attacks [4]. Another risk associated with cloud computing is derived directly from its distributed nature, which implies an increase in communications and network traffic as data must be transferred and synchronized across various infrastructure or between centralized systems and clients that consume their data. Therefore, associated risks are those incurred from data-in-transit, such as man-in-the-middle, side-channel, and spoofing attacks. Finally, malicious insiders can abuse their privileged access and launch attacks against residing data [4], resulting in a detrimental impact on the organization's trust and reputation that could potentially result in compliance fines imposed by governing bodies [7]. Although encryption has been widely used to secure data in the cloud, its algorithms are resource-intensive which creates additional overhead to the overall system performance [8-9].

Another common IT risk that transpired into cloud computing is identity management and access control of users. The most common authentication mechanisms consist of password and token-based systems. Passwords are well-known to be susceptible to various attacks, from brute force and interception to insecure or poorly configured storage, and thus stolen or being used in a fraudulent manner. Meanwhile, even though token-based authentication can be harder to reproduce or reuse when intercepted, such systems rely heavily on physical medium which can be easily transferrable to other users [8-10]. As a solution, biometric authentication enables individual identification through characteristics or traits. Therefore, biometric systems are characterized as reliable in applications that require higher levels of security such as border control [11]. However, its broader application in various fields have been hampered by many factors, including imperfect acquisition conditions and noisy data, as well as data tampering through fake biometrics which may result in an existing user being falsely accepted or rejected. Nevertheless, biometric authentication presents a significant advantage over other mechanisms, especially as modern systems like mobile devices are increasingly being launched with embedded biometric sensors.

The medical industry has not shied away from embracing the cloud paradigm. As such, the heightened use of new technologies has immensely changed the traditional practice of how healthcare providers handle and analyze patient data. Healthcare practitioners demand real-time access to patient data across multiple devices [12]. In particular, the deployment of modules that view, share, and store medical images offers an attractive solution. Since medical images are paramount tools for supporting diagnostics and treatment [13], medical institutions are highly keen on adopting cloud services. However, medical data and related applications are subject to rigorous legal as well as ethical regulations regarding data security [12]. Moreover, patients are increasingly voicing their concerns regarding privacy and valuable data being stored by third parties. Although cost-effective benefits attract healthcare institutions, a major concern revolves around the hosting of sensitive data, such as medical records and images, in public or hybrid clouds.

Hence, two topics related to cloud security in healthcare become relevant: protection of data residing in cloud servers and digital identity recognition. This paper will address the problems of digital identity and data security on public servers by proposing an architecture that uses a data fragmentation method to secure data and provide an authentication service that manages user identity through multimodal biometric modalities. By adopting biometric authentication with multiple modalities, inherited problems of using biometrics, such as noise in the data and imperfect acquisition conditions, are reduced so that the recognition accuracy is therefore enhanced. The suggested data storage strategy entails fragmenting medical data in the cloud and distributing it over several nodes of a distributed NoSQL database. It can also be combined with alternative data security mechanisms, such as encryption and anonymization. Moreover, the proposed method can be applied across multiple scenarios, especially where speed and confidentiality are of primary importance, such as in medical imaging and patient records, or DICOM (Digital Imaging and Communications in Medicine) files, an internationally standardized file type to store, transmit, retrieve, process, and display medical imaging along with related information. Medical data are arguably one of the most sensitive, vulnerable, and sought-after targets for bad actors, wherein tampering of any kind may ultimately result in loss of life [14]. However, healthcare institutions in general lack the required resources to address security concerns and thereby possess a fragile security posture. Therefore, the proposed solution not only enhances the overall organizational security, but also enables the system to be implemented in more niche settings with finite resources, such as remote diagnostics, telemedicine, or public cloud backups due to its lightweight nature.

This study aims to achieve the following objectives:

- To design a complete system for the use-case of medical image storage and to ensure the user digital identity.
- To evaluate efficient, lightweight solutions that ensure data security and organization in the cloud.

In Section 2, an overview of literature on data security in cloud and multimodal biometrics will be provided, followed by a detailed explanation of the system architecture and its individual components in Section 3. Finally, in Section 4, data security performance and biometric authentication methods will be evaluated, followed by a discussion on the identified benefits of the proposed system.

2 Literature Review

2.1 Data Security

By far, encryption is the most sought-after approach to

secure cloud data [1]. As addressed by [1], cryptographic systems must protect against both internal and external threats. However, most approaches only manage one or the other, particularly focusing on outsider threats. The works of [7] combined three different encryption algorithms and symmetric tokens to protect data residing in the cloud. Meanwhile, [15] proposed a system to protect sensitive data in the cloud by combining hashing and salting of symmetric and asymmetric encryption. Nevertheless, conventional encryption techniques increase computational overhead and inhibit the capability of performing queries on ciphered data. Therefore, [4] addresses the issue of encrypting against both insider and outsider threats with Attribute-Based Encryption (ABE) and Full Homomorphic Encryption (FHE). In addition, [14] studied the feasibility of using ABE to secure medical data on cloud systems. The design of the key for ABE increases in complexity, however, as attributes in the access strategy set mature, leading to increased operating expense in computing resources. Fully Homomorphic Encryption allows ciphertext to be analyzed, and when used in conjunction with other methods to deter outsider threats, demonstrates potential as a viable solution to protecting data in the cloud. However, FHE requires heavy computational expenditure [16-17].

Data anonymization has been suggested as a potential alternative for encryption to secure cloud data. Therefore, [18] compared various techniques using Map Reduce and identified its benefits and drawbacks. Also, [19] proposed a two-phase top-down specialization using K-Anonymity, taking advantage of the capabilities offered by Map Reduce to ensure data anonymization. On the other hand, [20] used semantic labeling to replace location coordinates with semantic categories.

Another technique that is recently experiencing a resurgence, with its origins dating back to the late seventies, is data fragmentation [21]. Its significantly lower computational demands highlight this method as potentially suitable for cloud environments. A range of fragmentation techniques were studied, analyzed, and classified into two distinct categories by [22]. Bahrami and Singhal [23] introduced a system which scrambles data using a pseudo-random permutation in a lightweight fashion based on a chaotic system. Kapusta and Memmi [22] also separated data into distinct groups based on their security level, which is directly proportional to the sensitivity of the data being stored. [24] stored data across different providers by fragmenting and categorizing. Meanwhile, [25] used Graphical Processing Units to implement fragmentation, encryption, and dispersion of data in the cloud. Furthermore, [26] extended the use of fragmentation to mobile phones for different policy strategies which need to access data from such devices.

2.2 Multimodal Biometric Authentication

Multimodal biometric authentication combines multiple biometric modalities to identify an individual [27]. It offers additional accuracy over unimodal biometric authentication as the fusion of multiple modalities overcomes the shortfalls of a single modality. [27] presents a method that uses Particle Swarm Optimization (PSO) to create a biometric image watermarking scheme. Meanwhile, [28] presents a system that combines face and fingerprint features using whale optimization and Maximally Stable External Regions, respectively, with a Support Vector Machine as a classifier. [29] proposes a biometric fusion system with face and fingerprint modalities using Arithmetic Light Adjustment to enhance the quality of images, Generic Algorithm to optimize the minutiae features, and Speed Up Robust Feature (SURF) to optimize facial features, with an Artificial Neural Network as a classifier. [30] combined face and fingerprint modalities using feature and decision-level fusion. They then extracted features using Scale Invariant Feature Transform (SIFT) and fed the combination of vectors through a K-Nearest Neighbor (K-NN), Support Vector Machine, Naïve Bayes (NB), and Radial based on Function classifiers. By combining the aforementioned modalities, [31] proposes a framework for securing and authenticating images in transit using face, fingerprint, and iris modalities with a Visual Sharing Neural Network (VSNN) along with Recurrent Neural Network-Bidirectional Long Short-Term Memory (RNN-BiLSTM) models to classify. The works of [32] fuse three unimodal biometric systems based on the face, fingerprint, and iris at a score-level based on Choquet Integral using PSO.

One of the main criticisms of biometric authentication is privacy risks, particularly the storage of biometric images or features in the template database. Governing bodies classify such data as sensitive and therefore ought to be protected in secure systems. Given this scenario, [33] proposes a watermarking-based approach that protects biometrics by embedding an iris image into a face image. Also, [34] and [35] provided a comprehensive review of various multimodal biometric techniques, along with a detailed review of biometric template security mechanisms. In their MFA-MB system, [36] secures templates using arithmetic hashing of random projections of biometric data, combined with a key derived from the user password to produce efficient and renewable authentication templates. Meanwhile, [37] proposes transformation-based template protection by incorporating biometric cryptosystems with cancellable biometrics.

In considering the previous methods, this study aims to secure cloud data by proposing a system that combines the benefits offered by data fragmentation and multimodal biometrics. Moreover, the proposed system has a potential for extension to allocate more data types, especially those explicitly associated with specific scenarios such as embedded systems or the Internet of Things (IoT).

3 Design of the Proposed System

In this section, a system design that stores and secures data in the cloud using a fragmentation technique and manages the identity of users through multimodal biometrics, is introduced. As previously mentioned, the proposed system aims to provide an alternative for data security in cloud computing.

Although efficient at securing data, encryption systems add complexity and overhead. Therefore, the study has

opted to implement a security method that splits data into byte chunks before inserting it into a distributed database cloud environment. Given the nature of medical images, the study opted for a NoSQL database to store data due to its native capabilities of storing unstructured data. Regarding the authentication and identity management of users who will interact with the application, a multimodal biometric authentication mechanism was implemented based on face and fingerprint modalities. Biometric-based authentication focuses on sensing and detecting individual, inherent characteristics. Unlike other authentication methods, biometric traits cannot be guessed, forgotten, misplaced, or easily forged. Moreover, portable devices such as laptops, tablets, and smartphones are increasingly embedding stock biometric sensors as a native authentication mechanism for users. However, single modality methods are susceptible to noisydata due to imperfect acquisition conditions. For instance, face image quality may be greatly affected by illumination conditions or facial expressions. Therefore, the fusion of two or more modalities in multimodal biometrics is intended to overcome such limitations and increase the recognition rate of biometric systems as well as safeguard against fake biometrics.

3.1 System Architecture

The example cloud scenario described in this paper is a repository of medical data and images stored across different cloud providers where data are fragmented and stored into randomized chunks. Typical data include files in the DICOM format as well as single images in other image formats or text data. The platform is oriented toward both public and private cloud deployment models.

The architecture proposed in this paper, as shown in Figure 1, is divided into separate cloud providers where Provider A and Provider B will be a public cloud setting to store user data; the cost-effective and scalable characteristics of the public cloud closely match this use-case. Meanwhile, Provider C will store biometric templates that can either be deployed in a private cloud or through an on-premises setting. Additional control over the underlying infrastructure allows more security measures to be deployed for protecting biometric data. This implementation also uses the proposed random pattern fragmentation (RPF) algorithm to protect biometric samples in the NoSQL biometrics database as it provides a lightweight and secure alternative to other protection mechanisms. The client application will then receive biometric samples from the devices and create a model which will be used to recognize the user. Afterwards, biometric data is simultaneously captured and sent with the user touching the fingerprint scanner while looking at the camera. During the enrollment stage, this model will be stored in the NoSQL biometric database. At the authentication stage, the samples are matched in the Authentication Server (AS) using the distance between all pairs of corresponding images and taking their average. The resulting score is then combined, and a final decision is made.

Once authenticated, the user will be able to define the fragmentation algorithm parameters such as the size of each individual chunk and how many splitfiles will be used to store data. Then, the client device will establish the fragmentation pattern for each file individually which will deter attackers from attempting to replicate a fragmentation pattern on a different file in the event that a key pattern is compromised.

The AS will be positioned between the client device and biometric database, handling all communications. RPF is being applied to the biometric data on the AS where the fragmentation key is securely stored. For each interaction with the data, the AS will regenerate the pattern for the file before storing chunks in the database. This will deter attackers from re-using a pattern key if acquired.

A Virtual Private Network (VPN) connection is placed between the user and the system which allows protected communications between the client and the cloud from man-in-the-middle attacks. As shown in Figure 1, the journey begins with the collection of user biometric samples. Then, the client application will communicate with the AS to interact with the biometric template database. Once the user is authenticated, a VPN route is established, with which the user will interact, to access the database residing in the public cloud wherein the fragmented data is located.

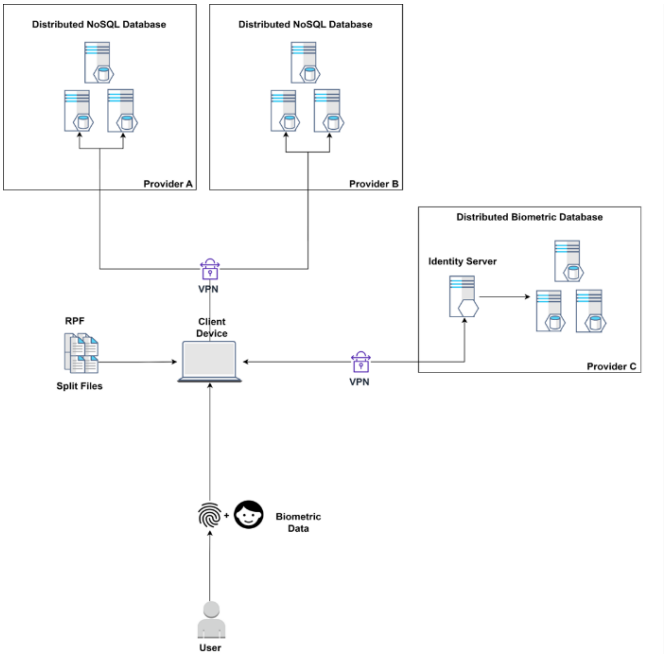

Figure 1 Architectural Overview of the Proposed System

Figure 2 highlights the flow of data between the different system components. Biometric data will first be transferred between the biometric and client device through a physical medium. From there, the biometric samples are sent from the client device to the AS over the internet through an encrypted connection. The AS then communicates with the distributed Biometric database to upload and retrieve samples. Once biometric authentication procedures are complete, the AS will provide an authentication token to the client device which will be used to interact with the RPF databases. Afterwards, split files containing randomized chunks will be sent from the client device.

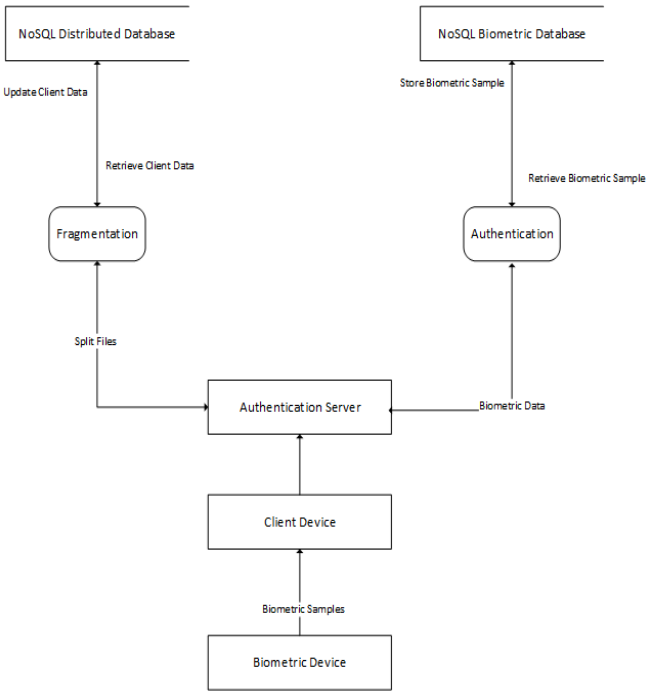

Figure 2 Data Flow Diagram of the Proposed System

3.2 Data Security

As previously mentioned, the chosen solution to ensure the security of data inhabiting the cloud infrastructure is to effectively store it in different randomized fragments, or chunks. This technique, extended from [38], consists of separating data into aleatory fragments before storing them in a distributed NoSQL database as described in [38]. If the database is compromised, an attacker would only be able to partially retrieve the data. The randomization algorithm is a cryptographically secure implementation which leverages the Operating System seed to generate randomness [39]. This implementation differs from others as its seed is harder to access due to Operating System restrictions, prohibiting two processes from simultaneously accessing the seed. Moreover, even if the perpetrator successfully acquires all the data, without the fragmentation pattern, attempts to reconstruct the data into its original form would be frustrated. Furthermore, in the unlikely event that an attacker determines the fragmentation pattern for a specific file, the remainder of the data would still be secure as a new pattern is generated for each file. This method is categorized as bitwise fragmentation with its potential applications ranging across different scenarios, in particular, those where speed and confidentiality are of the utmost importance such as in medical data storage, IoT, and backups. In an era where devices are ever portable and interconnected with limited resources, a lightweight mechanism to store data is paramount. As a result, the study opted to include a NoSQL database to provide a higher level of management over fragmented data. Moreover, leveraging the scalability and performances brought by the nature of NoSQL allows easier storage of unstructured data compared to its SQL counterparts.

The RPF is composed of:

Chunks: mutable bytes used to split the original file into an array of raw bytes wherein the user manually sets its value.

SplitFiles: a specialized structure used to store randomized chunks prior to their insertion into the database wherein the user can also manually set its value.

The RPF algorithm uses a randomized function which based on a permutation of N elements set by the user, calculates the pattern indexes. It contains two stages:

Disassembly: As summarized in Figure 3, the original file is partitioned into N chunks which are then inserted into splitfiles in a randomized method. The length of each split file is relative to the applied associated pattern which is set by the user. Unlike other similar approaches, the header is not segregated, which further hinders attempts to reconstruct the file.

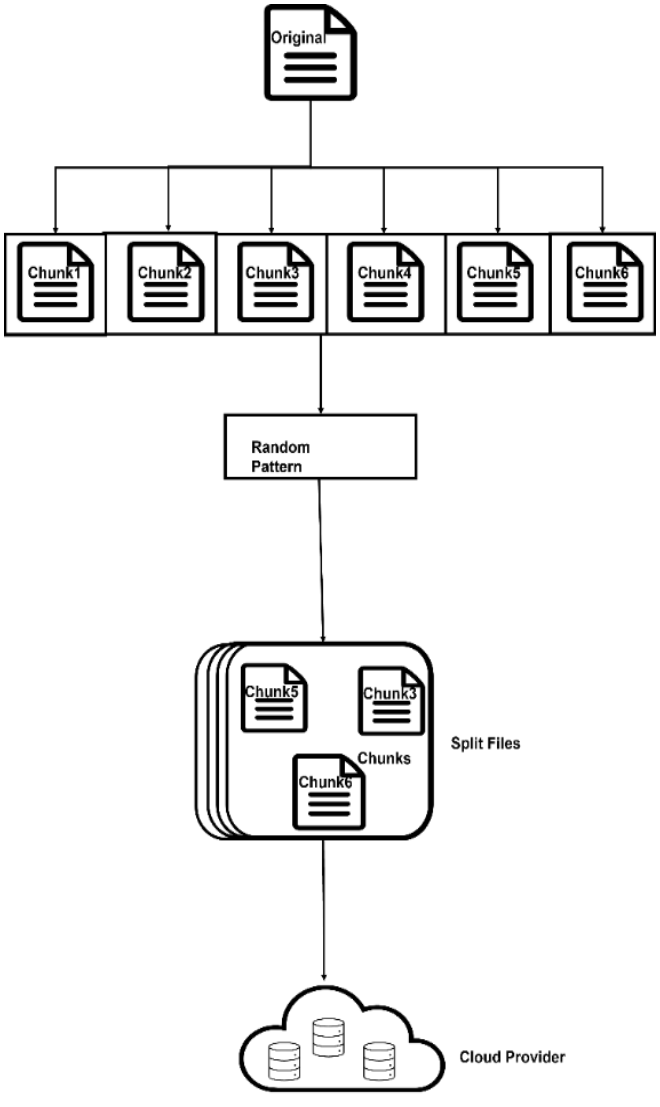

Figure 3 RPF disassembly stage.

Reassembly: As shown in Figure 4, the split file is recovered from the cloud, then opened on the client machine. Chunks residing in each split file are reorganized in a dictionary structure according to the pattern of the original

files. Once each chunk is regrouped into the original position, the reassembled file is stored in the client device.

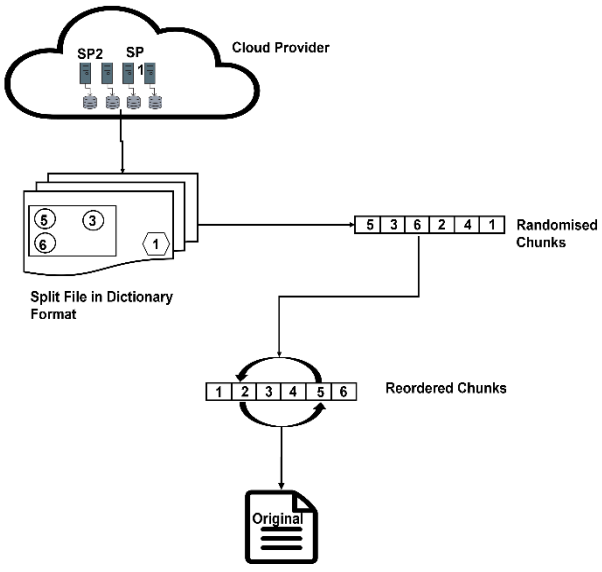

Figure 4 RPF in the reassembly stage.

As an additional layer of security, the proposed RPF method adds a Point-to-site TLS-based VPN connection between the client machine and cloud infrastructure. This aims to deter man-in-the-middle attacks against the solution as all communications between the client machine and cloud will be secured. Once connection is established and the user authenticated, the splitfiles will be concurrently sent to the distributed NoSQL database. During the reconstruction stage, upon authentication and connection to the VPN, the user will concurrently query the database for the desired files while the client device will reassemble and deserialize the chunks into their original form.

In particular, for the use-case of storing medical images, the proposed architecture is capable of managing multiple streams of medical data. The chosen RPF algorithm leverages concurrency paradigms to allow data to be transmitted in an asynchronous manner. Moreover, its lightweight design and dedicated support for unstructured data allow the proposed architecture to store various file formats such as DICOM.

3.3 Biometric Authentication

Based on the weighted sum of scores in each modality, the multimodal biometric authentication system implements a score-level fusion to combine face and fingerprint modalities. The fingerprint, one of the oldest and most used biometric modalities, is captured when its image is produced upon impression against a specialized sensor which is connected or embedded into a device. In the proposed implementation, the fingerprint is pre-processed while the Oriented Fast and Rotated Brief (ORB) [40] is used to extract feature points of the fingerprint with Hamming distance used for recognition. As shown in Figure 5, the minimal distance rule was used at the matching score-level where D represents the image while x_i and y_i correspond to the images being analyzed.

$$D_H = \sum_{i=1}^{k} |x_i - y_i|$$

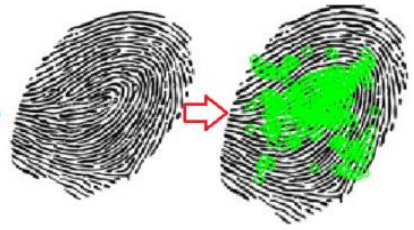

Figure 5 Fingerprint Recognition using ORB. Adapted from [41]

In the face implementation, multi-task cascaded convolutional neural networks (MTCNN) [42] are used for face detection and extracting facial coordinates. MTCNN (Figure 6) constitutes a three-stage multi-task deep CNN, where candidate windows are created through a fast Proposal Network (P-Net). Afterwards, candidates were refined in a refinement network (R-Net). Finally, the Output Network (O-Net) produces the final bounding box of the face.

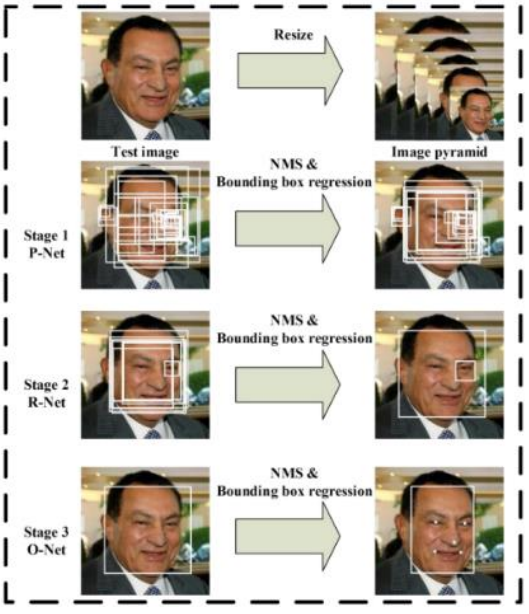

Figure 6 Implementation of MTCNN. Adapted from [42]

Meanwhile, FaceNet [44] was used to perform facial recognition. As shown in Figure 7, it consists of a face recognition pipeline which learns mapping from faces to a position in a multidimensional space where the distance between points directly correspond to a measure of face similarity. To determine such similarity, the Euclidean distance shown below was used, where x_i and y_i correspond to two vector points.

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

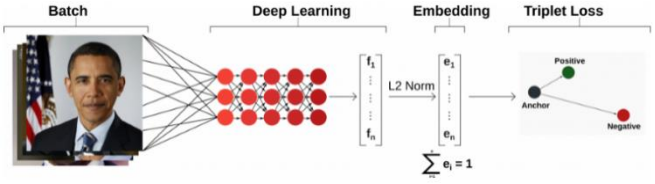

Figure 7 Face Recognition with FaceNet. Adapted from [45]

As shown in Figure 8, biometric samples were captured and sent to the AS. If a user is enrolling on the system, then samples are directly stored in the database. However, during authentication, the samples are compared against those stored in the database and their scores are then fused. Based on the threshold value of the combined scores, the decision module will enable either a successful or an unsuccessful attempt. As mentioned in 3.1, biometric templates stored in the database are protected using the RPF algorithm where features are separated into chunks and scrambled before being inserted into the Biometric NoSQL database.

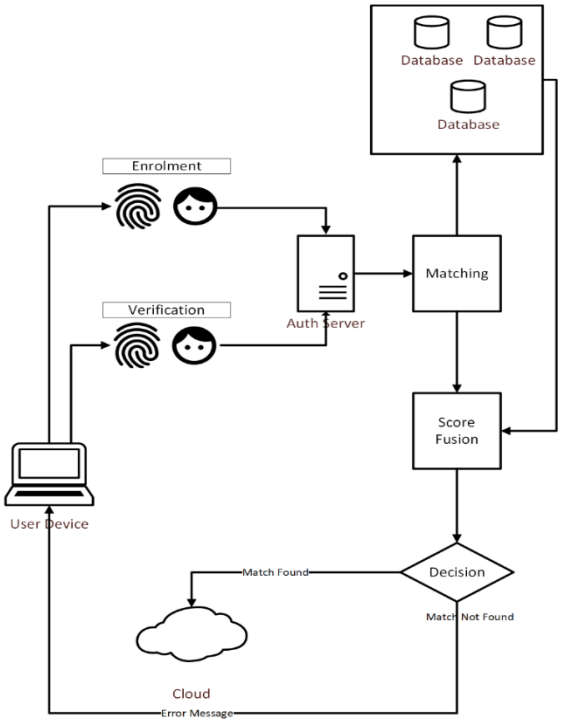

Figure 8 Overview of the Biometric Authentication Process

Finally, distance scores are normalized using the following function:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

where x corresponds to the distance scores. Once normalized, the scores are fused using a weighted sum where weights are experimentally attributed to each modality as per the formula:

$$\bar{x} = \sum_{i=1}^{n} w_i' x_i$$

4 Experiments and Results

4.1 Data Security

The RPF algorithm performance was tested using a dataset of five file types (.dcm; .jpeg; .pdf; .docx; and .png), each with 500 KB in size. Although the RPF algorithm allows the setting of chunk size and the number of splitfiles, for the purpose of the experiment, two splitfiles, with each containing 1000 byte chunks, were used. The experiment evaluated the total time (latency) of processing a file (fragmenting), uploading to the cloud, downloading from the cloud, and reconstructing to its original state.

For the experiment, Virtual Private Servers running Linux Ubuntu 20.04 LTS were launched across two different cloud providers, while a MongoDB database was installed on the servers in a distributed configuration. The implementation was compared against the works of [45] which used RPF in combination with Cassandra, a column-based NoSQL database. The Cassandra implementation followed a similar configuration as MongoDB with multiple servers distributed across two CSPs (Azure & AWS). An AMD Ryzen 7 5800H device with a 3.2 GHz CPU running Windows 11 was used, with a direct connection to a router and no other programs opened. All methods were tested under the same conditions, whereupon connecting to the cloud VPN, the files were separated into different chunks which are then scrambled and inserted into splitFiles. These splitFiles are concurrently sent to the database where they are permanently stored. Afterwards, the splitFiles are simultaneously downloaded while the chunks are removed, rearranged into their original positions, and the final byte array is used to reconstruct the original file which is then stored back into the client machine.

Many fragmentation methods have opted to store file headers separately from the remainder of the files. Given the latency impact of this action and the use-case this project targets, where speed is essential, the study decided against using this approach across all data types. Therefore, AES 256 was used to compare performance against the RPF algorithm due to its widespread popularity and use in cloud computing to secure data. The latency of encrypting, uploading, and downloading a file was measured in the AES 256 implementation, while for uploading and downloading the original file was through the SSH File Transfer Protocol (SFTP).

The results shown in Figure 9 highlight the superior performance of MongoDB, with an average latency of 5 milliseconds, over Cassandra whose average latency was 15 milliseconds; that is, MongoDB had the best performance across all systems. Given its bitwise nature, the RPF algorithm had a similar performance against all datatypes as it operates independently from file types and file associations.

The use of RPF plus the time required for its relative storage along with the use of AES and plaintext representation of the file without any modification is compared in Figure 10. It is worth noting that the MongoDB upload time and the implementation of the RPF method are considered. The performance effects of implementing encryption are evident. AES encryption took an average of 83 milliseconds while RPF had an average of 5 milliseconds. Notably, the single file upload took significantly more time than RPF with MongoDB as the RPF fragmentation algorithm allows for more data to be transferred simultaneously compared to the SFTP protocol.

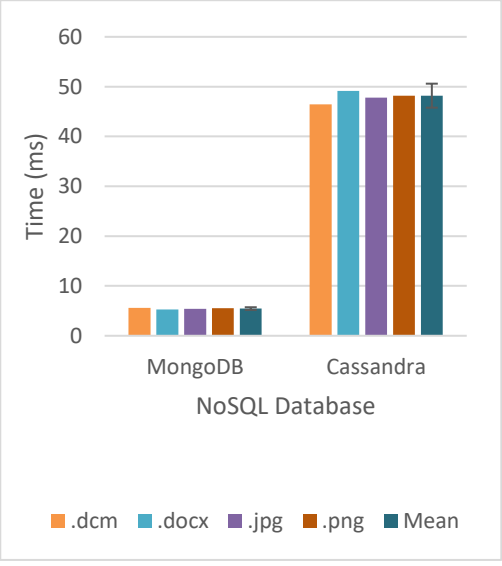

Figure 9 RPF method performance analysis across different NoSQL Databases. The total time describes RPF implementation, transfer time to the cloud, and storing into the database.

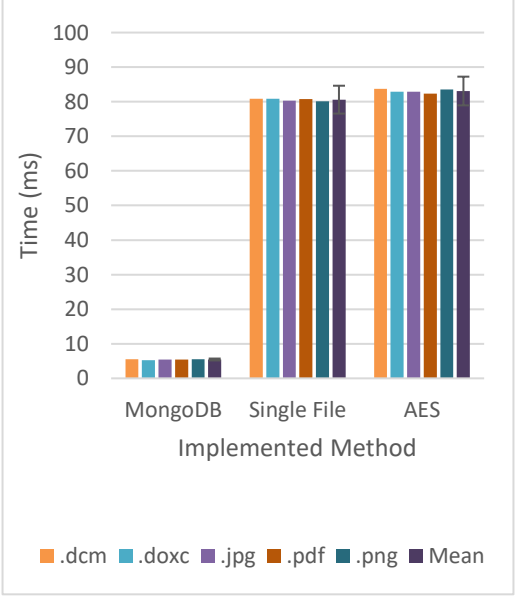

Figure 10 Performance comparison of sending the original file and an encrypted file by the Upload_V6_RPF

4.2 Multimodal Biometric Authentication

For the biometric authentication experiment, the BIOSECURE [46] database was used to measure the performance of the chosen algorithms. This database comprises multiple modalities to identify more than 100 users, among these, face and fingerprint. For the experiment, 50 random users were selected along with 12 samples of their face and fingerprint modalities.

In a biometric system, the Equal Error Rate (EER) is the value where the acceptance and rejection rates are equal. The study utilized a genuine score set and impostor score set for each modality. To calculate the False Acceptance Rate (FAR) and False Rejection Rate (FRR), the following formulas were utilized:

$$FAR(k) = \frac{100 * FRG(k)}{length(G(k))}$$

$$FRR(k) = \frac{100 * FAI(k)}{length(I(k))}$$

$$EER(k) = \frac{FAR(k) + FRR(k)}{2}$$

The system performance was evaluated for both modalities at different decision thresholds, ranging from 0.0 to 1.0. Figures 11–16 indicate that, at small threshold values, FAR remains very high while FRR is low. Meanwhile, as the threshold increases, the value of FAR decreases while FRR increases. For fingerprint authentication (Figure 11 and 12), the best threshold value is 0.4, whereas for face recognition (Figure 13 and 14), the best threshold value is 0.38. It is clear from the results that threshold values across both modalities are different. Figures 13 and 16 show that the EER value is deemed to be the intersection between FAR and FRR. The fingerprint and face systems had an EER of 0.39 and 0.37, respectively.

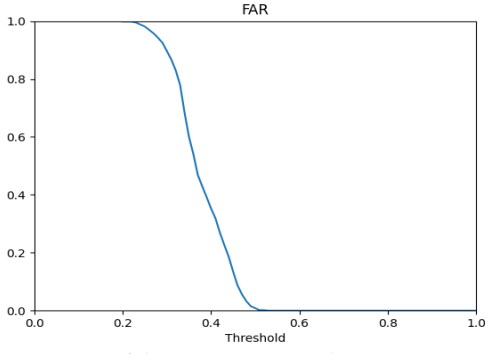

Figure 11 FAR of the Fingerprint Implementation

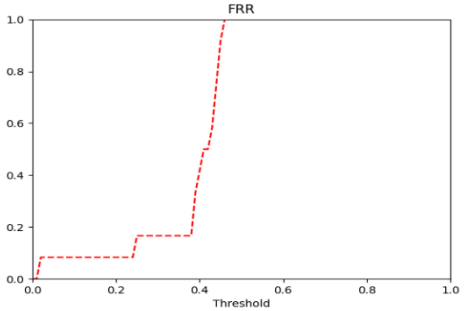

Figure 12 FRR of the Fingerprint Implementation

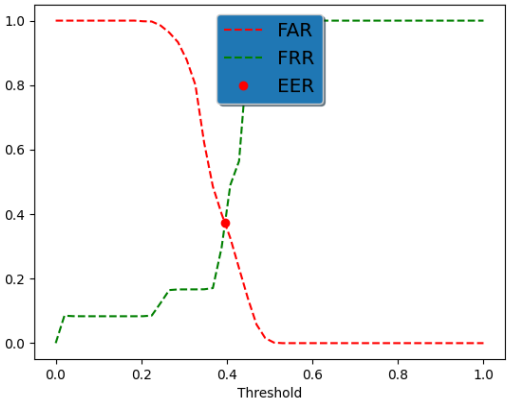

Figure 13 EER of the Fingerprint Implementation

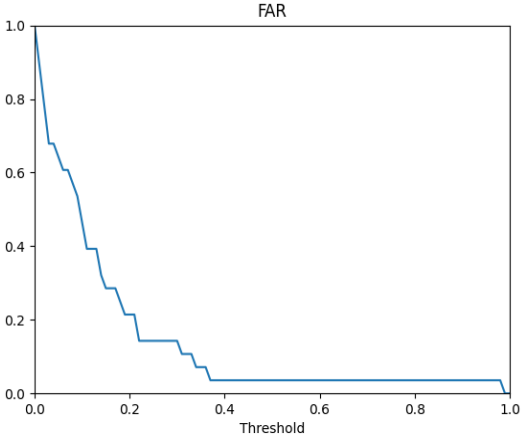

Figure 14 FAR of the Face Implementation

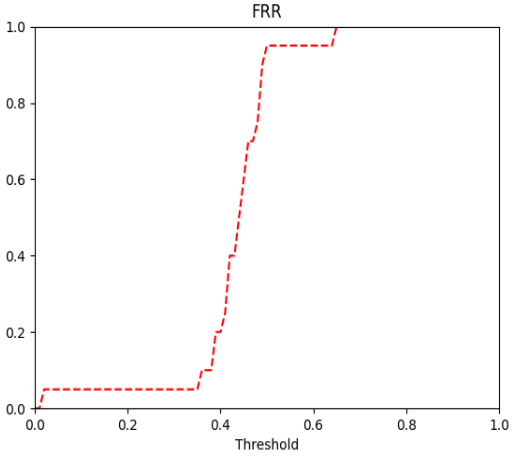

Figure 15 FRR of Face Implementation

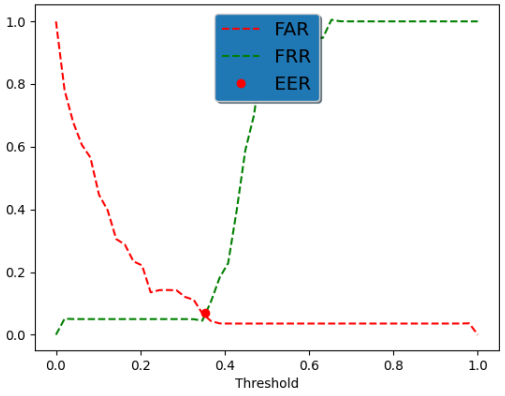

Figure 16 EER of Face Implementation

Figures 17 and 18 show the resulting FAR and FRR when using score-level fusion. The FRR threshold is 0.25, showing an improvement when compared to single modalities, whereas the FAR threshold is 0.45. Meanwhile, the fusion EER is 0.39 as shown in Figure 19.

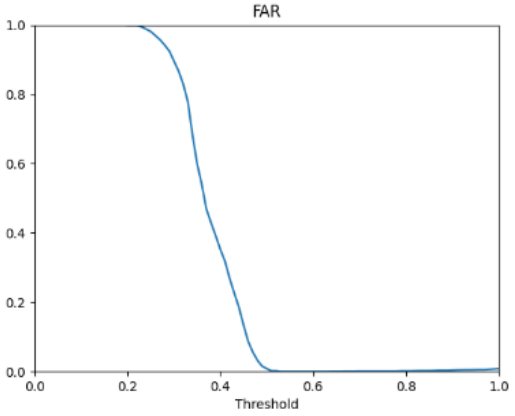

Figure 17 FAR of the Multimodal Implementation

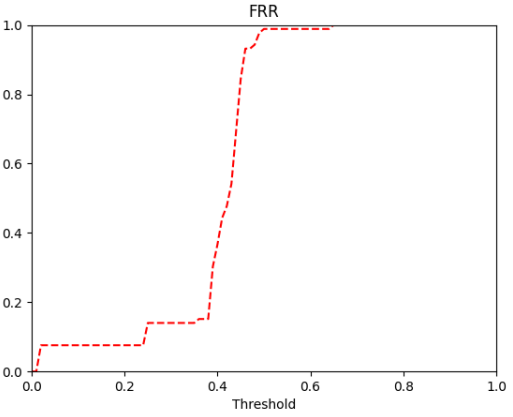

Figure 18 FRR of the Multimodal Implementation

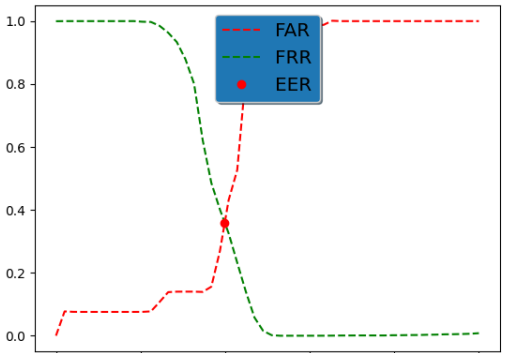

Figure 19 EER of the Multimodal Implementation

5 Conclusion

Despite its popularity and adoption across various industries, data security as well as the privacy of users in cloud computing remain priority concerns against the adoption of this paradigm. Traditional security mechanisms, such as encryption, have been unable to adequately protect cloud data due to its high resource requirements and unstructured nature. Therefore, the risks are higher for medical data, as any breach in its confidentiality, integrity, and availability may result in loss of life. Therefore, this study proposed a system that securely stores data in a cloud using RPF and a NoSQL database to leverage its unstructured nature and facilitate the management of a wider range of data types. Moreover, a multimodal biometric authentication system based on the face and fingerprint modalities to allow users access to the cloud environment was introduced.

Results indicate a strong processing speed for all NoSQL databases implemented by RPF with a particular highlight of MongoDB, demonstrating fast and secure capabilities when compared to AES encryption. Thus, the proposed system can be used to protect data across multiple scenarios. Particularly in healthcare, it can be used for general repositories of medical data or as a standalone method for private use. For further applications within medical scenarios, the integration of more commonly used industry methods, such as PACS (Picture Archiving and Communication System, can be employed. Also, the flexibility demonstrated by the proposed system allows the RPF method to be combined with other mechanisms, such as encryption for scenarios with higher security requirements. Furthermore, the data-agnostic and vendor-agnostic approach taken to design this system enhances its potential to be trialled in other industries or scenarios, such as big data, or backup cloud images (e.g., Dropbox).

References

[1] Z. Yan, R. Deng, and V. Varadharajan, "Cryptography and Data Security in Cloud Computing", *Information Sciences*, vol. 387, pp. 53-55, 2017.

[2] P. Mell and T. Grance, "The NIST definition of cloud computing", 2011.

[3] M. Rady, T. Abdelkader and R. Ismail, "Integrity and Confidentiality in Cloud Outsourced Data", *Ain Shams Engineering Journal*, vol. 10, no. 2, pp. 275-285, 2019.

[4] X. Sun, "Critical Security Issues in Cloud Computing: A Survey", *2018 IEEE 4th International Conference on 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)*, 2018.

[5] M. G. Avram, "Advantages and Challenges of Adopting Cloud Computing from an Enterprise Perspective", *Procedia Technology*, vol. 12, pp. 529-534, 2014.

[6] Pallathadka, G. Sajja, K. Phasinam, M. Ritonga, M. Naved, R. Bansal and J. Quiñonez-Choquecota, "An investigation of various applications and related challenges in cloud computing", *Materials Today: Proceedings*, 2021.

[7] N. Dahiya, "implementing multilevel data security in cloud computing", *International Journal of Advanced Research in Computer Science*, vol. 8, no. 8, pp. 146-152, 2017.

[8] S. Aldossary and W. Allen, "Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions", *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 4, 2016.

[9] P. Kumar, P. Raj and P. Jelciana, "Exploring Data Security Issues and Solutions in Cloud Computing", *Procedia Computer Science*, vol. 125, pp. 691-697, 2018.

[10] R. Rao and K. Selvamani, "Data Security Challenges and Its Solutions in Cloud Computing", *Procedia Computer Science*, vol. 48, pp. 204-209, 2015.

[11] G. Masala, P. Ruiu and E. Grosso, "Biometric Authentication and Data Security in Cloud Computing", in *Computer and Network Security Essentials*, K. Daimi, Ed. Springer, Charm, 2018.

[12] J. Vincent, W. Pan and G. Coatrieux, "Privacy protection and security in the eHealth cloud platform for medical image sharing", *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2016.

[13] M. Marwan, A. Kartit and H. Ouahmane, "Using cloud solution for medical image processing: Issues and implementation efforts", *2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech)*, 2017.

[14] S. Chandel, T. Ni and G. Yang, "Enterprise Cloud: Its Growth & Security Challenges in China", *2018 5th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2018 4th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, 2018.

[15] A. Arora, A. Khanna, A. Rastogi and A. Agarwal, "Cloud security ecosystem for data security and privacy", *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, 2017.

[16] N. Samardzic et al., "F1: A Fast and Programmable Accelerator for Fully Homomorphic Encryption", *MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture*, 2021.

[17] C. Gouert, D. Mouris and N. Tsoutsos, "New Insights into Fully Homomorphic Encryption Libraries via Standardized Benchmarks", *Cryptology ePrint Archive*, vol. 2022, no. 425, 2022.

[18] E. Stefanov, C. Papamanthou and E. Shi, "Practical Dynamic Searchable Encryption with Small Leakage", *Network and Distributed System Security Symposium*, 2014

[19] R. Sreedhar and D. Umamaheshwari, "Big-Data Processing with Privacy Preserving Map-Reduce Cloud", *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 3, no. 1, pp. 343-350, 2014.

[20] Z. Priyanka, K. Nagaraju and Y. Venkateswarlu, "Data

Anonymization Using Map Reduce on Cloud based A Scalable Two-Phase Top-Down Specialization", *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 2, no. 12, pp. 3879-3883, 2014.

[21] O. Barak, G. Cohen and E. Toch, "Anonymizing mobility data using semantic cloaking", *Pervasive and Mobile Computing*, vol. 28, pp. 102-112, 2016.

[22] A. Shamir, "How to share a secret", *Communications of the ACM*, vol. 22, no. 11, pp. 612-613, 1979.

[23] K. Kapusta and G. Memmi, "Kapusta, Katarzyna, and Gerard Memmi. "Data protection by means of fragmentation in various different distributed storage systems-a survey", *arXiv*, vol. 170605960, 2017.

[24] M. Bahrami and M. Singhal, "A Light-Weight Permutation Based Method for Data Privacy in Mobile Cloud Computing", *2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*, 2015.

[25] K. Kapusta and G. Memmi, "Data protection by means of fragmentation in distributed storage systems", in *2015 International Conference on Protocol Engineering (ICPE) and International Conference on New Technologies of Distributed Systems (NTDS)*, Paris, pp. 1-8, 2015.

[26] G. Memmi, K. Kapusta and H. Qiu, "Data protection: Combining fragmentation, encryption, and dispersion", in *2015 International Conference on Cyber Security of Smart Cities, Industrial Control System and Communications (SSIC)*, Shanghai, pp. 1-9, 2015.

[27] H. Dev, T. Sen, M. Basak and M. Ali, "An Approach to Protect the Privacy of Cloud Data from Data Mining Based Attacks", in *2012 SC Companion: High Performance Computing, Networking Storage and Analysis*, Salt Lake City, pp. 1106-1115, 2012.

[28] P. Bedi, R. Bansal and P. Sehgal, "Multimodal Biometric Authentication using PSO based Watermarking", *Procedia Technology*, vol. 4, pp. 612-618, 2012.

[29] T. Kumar, S. Bhushan and S. Jangra, "An Improved Biometric Fusion System of Fingerprint and Face using Whale Optimization", *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, 2021.

[30] T. Kumar, S. Bhushan and S. Jangra, "An Improved Biometric Fusion System Based on Fingerprint and Face using Optimized Artificial Neural Network", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 11, pp. 1568-1575, 2019.

[31] B. Somashekhar and Y. Nijagunarya, "FACE AND FINGERPRINT FUSION SYSTEM FOR IDENTITY AUTHENTICATION USING FUSION CLASSIFIERS", *International Journal of Computer Science & Engineering Survey (IJCSES)*, vol. 9, no. 123, 2018.

[32] M. Gayathri and C. Malathy, "Novel framework for multimodal biometric image authentication using visual share neural network", *Pattern Recognition Letters*, vol. 152, pp. 1-9, 2021.

[33] L. Haddada and N. Ben Amara, "Score-Level Fusion of Fingerprint, Face and Iris based on Choquet Integral", *2019 5th International Conference on Nanotechnology for Instrumentation and Measurement (NanofIM)*, 2019.

[34] C. Kant and S. Chaudhary, "A Watermarking Based Approach for Protection of Templates in Multimodal Biometric System", *Procedia Computer Science*, vol. 167, pp. 932-941, 2020.

[35] S. Choudhary and A. Naik, "Multimodal Biometric Authentication with Secured Templates — A Review", *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019.

[36] A. Mansour, M. Sadik and E. Sabir, "Multi-factor authentication based on multimodal biometrics (MFA-MB) for Cloud Computing", *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, 2015.

[37] M. Lee, A. Teoh, A. Uhl, S. Liang and Z. Jin, "A Tokenless Cancellable Scheme for Multimodal Biometric Systems", *Computers & Security*, vol. 108, p. 102350, 2021.

[38] N. Santos, W. Younis, B. Ghita and G. Masala, "Enhancing Medical Data Security on Public Cloud", *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*, 2021

[39] Python.org. "Secrets — Generate secure random numbers for managing secrets — Python 3.10.6 documentation", 2022.

[40] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, "ORB: An efficient alternative to SIFT or SURF", *2011 International Conference on Computer Vision*, 2011.

[41] T. Zh. Mazakov, Sh. A. Jomartova, T. S. Shormanov, G. Z. Ziyatbekova, B. S. Amirkhanov and P. Kisala, "THE IMAGE PROCESSING ALGORITHMS FOR BIOMETRIC IDENTIFICATION BY FINGERPRINTS", *News of the Academy of Sciences of the Republic of Kazakhstan*, 2022.

[42] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks", *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499-1503, 2016.

[43] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering", *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[44] N. Santos, B. Ghita and G. Masala, "Enhancing Data Security in Cloud using Random Pattern Fragmentation and a Distributed NoSQL Database", *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019

[45] P. Madio, "A FaceNet-Style Approach to Facial Recognition", *Medium*, 2022. [Online]. Available: https://towardsdatascience.com/a-facenet-style-approach-to-facial-recognition-dc0944efe8d1. [Accessed: 05- Sep- 2022].

[46] J. Ortega-Garcia et al., "The Multiscenario Multienvironment BioSecure Multimodal Database (BMDB)", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1097-1111, 2010. Available: 10.1109/tpami.2009.76 [Accessed 5 September 2022].

Nelson Santos is a PhD. Candidate in Applied Computing at the University of Plymouth. The thesis title is "Data Security on Cloud Using Machine Learning." Before that, he received a BSc (Hons) in Computer and Information Security (2017) at the University of Plymouth. He is also a Cybersecurity Consultant at a UK-based company. His research is focused on data security, cloud security and multimodal biometrics. He has four publications, where one has won an award for best research paper at the 2018 International Conference on Intelligent Interactive Multimedia Systems and Services. Currently, he is an IEEE Student Member.

Bogdan Ghita received a PhD degree from the University of Plymouth, U.K., in 2005. He is currently an Associate Professor at the University of Plymouth and leads the networking area within the Centre for Security, Communications, and Network research. His research interests include computer networking and security, with a focus on network security, performance modelling and optimization, and wireless/mobile networking. He has published over 150 articles, has graduated 20 PhD students, and has been a principal investigator in a number of industry-led, national, and EU research projects in these areas. He was a TPC member for over one hundred international conference events and a Reviewer of the IEEE Communications Letters, Computer Communications, and Future Generation Computer Systems journals.

Giovanni L. Masala received a PhD degree in Applied Physics. He is a Senior Lecturer of Computer Science and Co-Lead of Future Human Signature Research Theme at the University of Kent, Canterbury, U.K. He is also a Visiting Research Fellow with the University of Plymouth, Plymouth, U.K. He has produced more than 90 publications in international journals and conference proceedings. He is involved with numerous international research grants and has been a principal investigator in a number of industry-led, national, international projects on such topics. His research interests are in artificial intelligence (AI) and robotics, AI in medical applications, cloud systems and data security. Dr Masala has been part of program committees and has chaired several international workshops and conferences in networks and security. He is a Guest Associate Editor in many journals on AI and an Associate Editor at Frontiers, in robotics and AI–computational intelligence.