Baraton Interdisciplinary Research Journal (2023), 9(Special Issue), pp 64-72
www.birj.ueab.ac.ke
ISSN:2789-3995
(Online)

STORAGE OF BIOMETRIC DATA IN MATRIX FORM (DATA THROUGH THE LENS OF SECURITY)

Vincent Agini*, Meshack Omagwa, Daniel Wagara, Stella Lasiti, Jade Abuga & Paul Francis
University of Eastern Africa, Baraton, P. O. Box 2500-30100, Eldoret, Kenya
*Corresponding author Email address: aginivincent254@gmail.com

Abstract

In contemporary times, individuals increasingly favor authenticating their biometric details through unique technological identification systems. Within this digital regime, concerns regarding data security have intensified, as users seek assurance of privacy amidst the proliferation of malicious activities such as data breaches. The primary aim of this study is to safeguard confidential information by employing the Hamming code technique as a noise mask against attacks, complemented by the utilization of steganography to conceal data, thereby mitigating the risk of permanent identity loss. Initially, a binary confidential matrix is constructed to identify certain locations within the matrix, ensuring data integrity. Subsequently, the outcome of a collision between two hashed templates is authenticated based on the final confidence score. The proliferation of various fingerprint schemes presents an additional challenge, necessitating the determination and selection of the most suitable scheme. Our proposed model, utilizing computer simulations of the system's design, enhances the security of the entire scheme. In the event of an attacker attempting to breach the reference template to access confidential biological information, our model ensures that the final confidential information remains safeguarded, impervious to leakage..

INTRODUCTION

In the wider deployment of user authentication in identity management systems there are increasing concerns about the security and privacy of the biometric details. Technological advancement has increased vulnerability within the data systems which successively may result in inconveniences in data storage.

Password techniques are the subject of intense debate in academia, and are arguably manifested to be often poor (Uludag, 2004) ; (Rathgeb, 2011)Passwords are not secure and never have been since almost immediately after passwords were invented, the first breach occurred. The module of passwords has some drawbacks since, upon password entry by the user, the system compares the entered password to the password file entry for the corresponding user Identification card; thus, a sound user code and password don't necessarily prove the real user's identity and most reliable are the techniques supported with biometric data (Tudor, 2001);(J. Menezes, 1996)

According to (Tibor, 2015) the word Biometry stems from Greek and has two phrases "bios and metrein‟ where „bios" means life and „metrein" means to measure. Biometrics involves identifying the topic which is supported by its special physical characteristic. The methodologies employed in biometrics include fingerprinting,

ISSN:2789-3995
(Online)

signature recognition, Keyboard Dynamics, Voice Recognition, and Hand Geometry

Biometric identifiers can relieve people from having to recollect difficult credentials or carry and protect tokens since the system recognizes a person by matching the individual's biological data with the digital files stored within the system database.

Csaba (Otti, 2015) noted that the primary commercially available hand geometry identification system was launched in 1974 which was the primary truly automatized system that would facilitate identification to access control and attendance tracking. This technique was later deployed in 2008 at Paks Nuclear Energy Plant in Hungary. The matcher module also encapsulates a decision-making module, within which a user's claimed identity is confirmed or a user's identity is established and supported by the matching score (Jain, 2004)

(Manual, 2008) Confirms that within the late 19th century, Sir Francis Galton, a scientist worked on a close study of fingerprints and he recorded that the chances of two individual fingerprints being identical were 1 in 64 billion. He identified the characteristics by which fingerprints may be identified (minutia), which are discontinuities that disrupt the flow of fingerprint ridges. Fingerprint verification with its phases, has been widely accepted as a key biometric recognizing technique in many applications like healthcare, enforcement, data system security, and physical access control among others.

According to (Wayman, Jain, Maltoni, & Maio, 2005) several challenges remain unsolved in designing a completely automatic and reliable fingerprint individualization system, especially when fingerprint images are of poor quality. Although automatic systems have improved significantly, the design of automated systems does not yet match the complex decision-making of a well-trained fingerprint expert as decisions are made to match individual fingerprints.

In addition, there have been some instances where fingerprint information is hacked for example, according to (Chi, 2016) BBC, "Not period ago, 70 million citizens' fingerprints and passport information was hacked within the Philippines." The Bangladesh heist also sheds light on the risks of rapid digitalization. Noting that a wise Biometric National identification (NID) will soon be assigned to Bangladesh Citizens, one can conclude that, the problem of biometric data in Bangladesh has not been resolved.

According to (Sam, 2021), "Lawyer Dudley Ochiel said that data protection impact assessment may be necessary for alleviating this gap. To mitigate these risks related to data loss, fraud, and hacking among others, we applied matrices to embed data to increase data security and efficiency.

Thus, in this research, we propose a data-hiding technique that will not only hide the data but also assure consumers of their privacy. The research will thus focus on the hamming code and steganography techniques for efficient data hiding:

According to (Hoffman, 1991), coding theory is "the study of methods for efficient and accurate transfer of information from one place to another". In our case, we are going to study and efficiently apply this method to store biometric data accurately. (Moreira, 2006), grouped codes into different types:

Baraton Interdisciplinary Research Journal (2023), 9(Special Issue), pp 64-72
www.birj.ueab.ac.ke
ISSN:2789-3995
(Online)

nonsystematic codes and systematic codes in which from the data bits k are derived redundancy bits (r=n-km), where n is the length of the code word. The code word consists of the data bits followed by the redundancy bits. (Gao, 2007) Also grouped codes into two groups;

Source coding (entropy coding): is the process of compressing source data to gain higher transmission effectiveness.

Channel coding (forward error correction): is the process of adding redundant bits to the message to ensure resistance against communication noise.

For our case, we shall use channel coding or forward error correction codes (Hamming Codes) that will detect and correct any errors that might occur on the stored data.

Steganography is a technique used for hiding data by generally embedding the secret message into another innocent-looking message. In steganography, the hidden message is important while the cover is not. (Schneier, 1999) Pointed out that a biometrics-based verification system works properly only if the verifier system can guarantee that the biometric data came from a legitimate person at the time of enrolment.

This study will be of great significance to companies, organizations, institutions, and firms to secure stored information from hackers, cyberpunk, geeks, netizen, and other skilled computer programmers. We will design and develop a computer simulation that is a reliable approach to creating a constant and non-intrusive transparent biometric technique that is therefore probative.

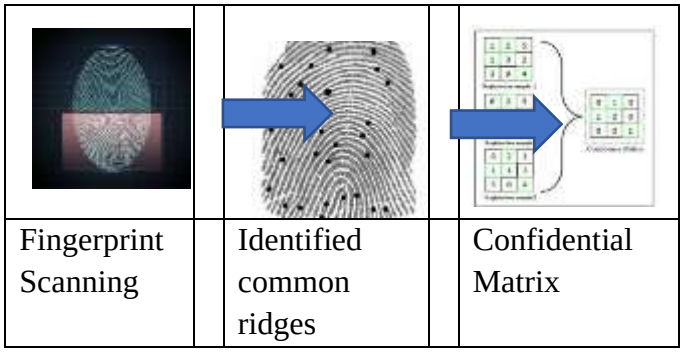
The study may be of importance also to researchers who wish to increase a similar study within the field of studies.

METHODOLOGY

Fingerprinting

Generally, methods for extracting fingerprint features can be classified into two primary categories. The first category is the ridges pattern-based feature extraction, and the second is the minutiae-based feature extraction. Several methods have been developed to extract useful information from fingerprint ridges. (Neil, 2004) Theorized that some of these methods examine the frequency and orientation of the ridges, while others develop mathematical models to represent the structure of the ridges.

Figure 1: Fingerprint Scanning



Fingerprint Scanning	Identified common ridges	Confidential Matrix

Hamming Code

Richard Hamming developed this method and it employs redundant bits within the message, these redundant bits are extra bits that are generated and inserted at specific positions in the message itself to enable error detection and correction. Usually, a 7-bit Hamming code is used. Once the redundant bits are embedded within the message, then the message is ready to be stored.

Baraton Interdisciplinary Research Journal (2023), 9(Special Issue), pp 64-72
www.birj.ueab.ac.ke
ISSN:2789-3995
(Online)

When storing data bits over a computer network, they are subject to getting corrupted due to interference and network problems, the corrupted data leads to errors. To mitigate these errors, we deploy hamming code, a block capable of detecting and correcting an error in our confidential matrix before concealing it.

Encoding a Message by Hamming Code

Step 1: Calculation the number of redundant bits.

If the message contains mm number of data bits, rr number of redundant bits are added to it so that mr can indicate at least (m + r+ 1) different states. Here, (m + r) indicates the location of an error in each of the $(m + r)$ bit positions, and one additional state indicates no error. Since rr bits can indicate 2rr states, 2rr must be at least equal to (m + r + 1). Thus, the following equation should hold $2r \geq m+r+1$.

Step 2: Position the redundant bits

The r redundant bits are placed at bit positions of powers of 2, i.e., 1, 2, 4, 8, 16, etc.

Step 3:Calculate the values of each redundant bit

A parity bit is an extra bit that makes the number of 1s either even or odd. The two types of parity are;

Even Parity − Here, the total number of bits in the message is made even.

Odd Parity − Here, the number of bits in the message is odd.

Each redundant bit, r_i, is calculated as the parity, generally even parity that is 2_n, where {n=0, 1, 2, 3……, n} based upon its bit position. It covers all bit positions whose binary representation includes a 1 in the ith position except the position of r_i. Thus; r1 is the parity bit for all data bits in positions whose binary representation includes a 1 in the least significant position excluding 1 (3, 5, 7, 9, 11, and so on), r_2 is the parity bit for all data bits in positions whose binary representation includes a 1 in the position 2 from right except 2 (3, 6, 7, 10, 11 and so on), r_3 is the parity bit for all data bits in positions whose binary representation includes a 1 in the position 3 from right except 4 (5-7, 12-15, 20-23 and so on). For a 7-bit hamming code we will have 4 data bits and 3 parity bits.

Using the formula 2n we calculate the parity bits positions as follows;

When n=0,

n=1,

n=2,

n=3,

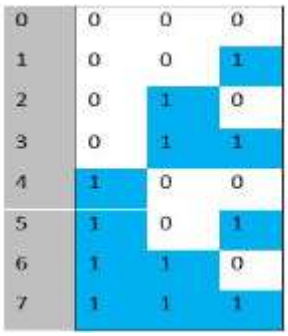


Figure 2: Truth Table

Since we are dealing with a 7-bit we will not include the last bit, if we were dealing with more bits, we were to include the last parity. According to Hamming, the parity bits are associated with;

Positioning of parity bits

Baraton Interdisciplinary Research Journal (2023), 9(Special Issue), pp 64-72
www.birj.ueab.ac.ke
ISSN:2789-3995
(Online)

p1=D3 D5 D7

p2=D3 D6 D7

p3= D5 D6 D7

Example

Generate hamming code for the message 0 1 0 1.

We know that $2p \geq p + m + 1$,
where p-parity bits and m-message or data

$2p \geq p + 4 + 1$

$2p \geq p + 5$

$p = 3 \Leftrightarrow 8 = 8$

Hamming Code - Error Detection and Error Correction

Let's understand the Hamming code concept with an example:

1	2	3	4	5	6	7
P1	P2	M1	P3	M2	M3	M4
0	1	0	0	0	0	1

Figure 3: Corrupted bits

Step 1:

For checking parity bit P3, from the parity check matrix, it depends on positions 4, 5,6&7. This way we will have the following bits, (0 0 0 1)

Step 2:

While checking for P2, again from the parity check matrix, it depends on positions 2, 3,6&7. (1 0 0 1)

Step 3:

According to the parity check matrix, checking for P1 depends on positions 1, 3,5&7. (0 0 0 1)

While checking the parity, if the total number of 1's is odd then write the value of parity bit P1(or P2 etc.) as 1 (which means the error is there), and if it is even then the value of parity bit is 0 (which means no error).

Hamming Code: Error Correction

To correct the errors, use the following steps:

Now the error word E will be in the form E3E2E1=101

Now we have to determine the decimal value of this error word 101 we get E = 5, which states that the error is in the fifth data bit. To correct it, just flip the fifth data bit.

So, the correct data will be:

1	2	3	4	5	6	7
P1	P2	M1	P3	M2	M3	M4
0	1	0	0	1	0	1

Figure 4: Corrected bits

Steganography

With steganography, the existence of a message is unknown. If it becomes known, it (mostly) can be read.

The red rose whispers of passion,

And the white rose breathes love,

O, the red rose is a falcon,

And the white rose is a dove.

But I send you a cream-white rosebud

With a flush on its petal tips,

Baraton Interdisciplinary Research Journal (2023), 9(Special Issue), pp 64-72
www.birj.ueab.ac.ke
ISSN:2789-3995
(Online)

For the love that is purest and sweetest,

by (John, 2022) Anything secret about this?

If there is any extra white space, it is encoded as 1 otherwise, it is encoded as 0.

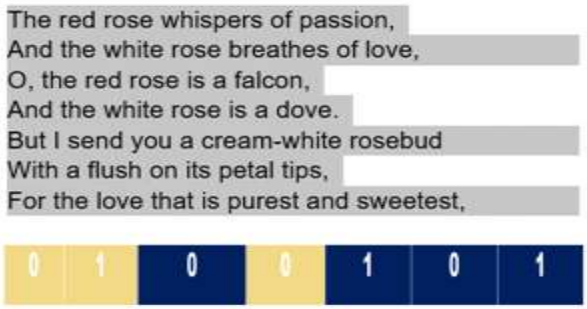


Figure 5: Embedding by the use of White Spaces

From the stanza above, a confidential matrix is generated by the white spaces.

The encoded text is invisible to human eyes in typical text editors (white spaces).

ANALYSIS

Authentication systems typically perform the following operations: -

Data capture – A sensor of some kind captures data from the biometric identifier used.

Enrollment – The captured data is analyzed and its unique features are stored as a digital template.

Authentication – When the enrolled user wants to authenticate themselves, their

Data Capture - Biometric data is thus captured once again and compared against the template generated by the enrollment.

Matching – An algorithm is used to compare whether there is a match between the stored template. If there is a match the user has been authenticated, otherwise access is denied.

Figure 6: Fingerprint visualization and scanning

The figure above shows a finger placed on the fingerprint scanner ready for scanning. The binary confidential matrix is first formed, which will identify the certainty locations in the matrix, verify the outcome of a collision between two hashed templates and authenticate based on the final confidence score.

After scanning, the scheme will cross-match every element within the selected hashed registration samples. The main purpose of the confidence matrix is to identify hashed bits, which can be categorized as confidence bits. When all the paired registration samples give the same value in a particular location, a matched collision is fulfilled and defined as the confidence bit location. For instance, if the same value is found in all the hashed registration samples at the same respective

Baraton Interdisciplinary Research Journal (2023), 9(Special Issue), pp 64-72
www.birj.ueab.ac.ke
ISSN:2789-3995
(Online)

location (x, y), the value "1" will be assigned to that confidence location (x, y)

If this condition is not fulfilled, the particular bit location will be labeled as "0" under the "no confidence" location. Finally, a binary confidence matrix will be generated. A successful match during authentication means that identity has been verified.

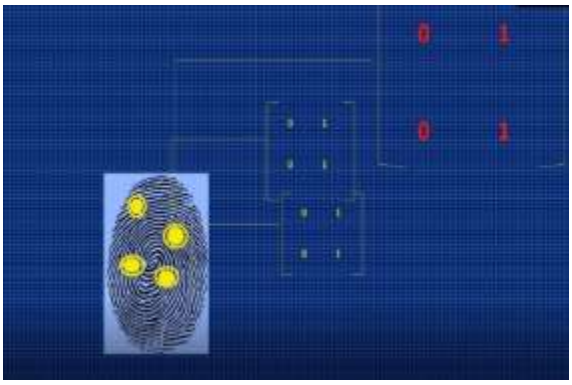


Figure 7: Final binary confidential matrix

After a successful collision between the hashed templates then, the final confidential matrix is formed which contains the biological features of the user.

Fingerprint verification

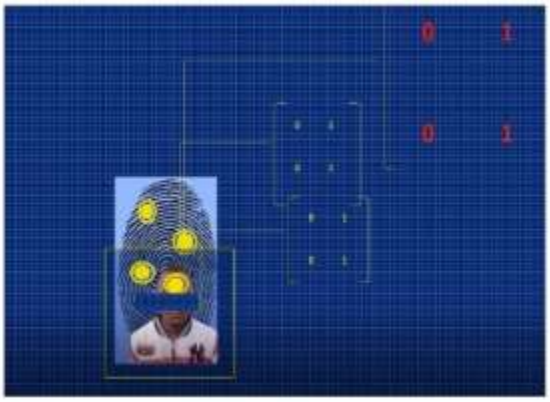


Figure 8: Fingerprint verification with image-sized passport photo

During registration, the system also takes an image-sized passport photo of the user. Thus, when verifying if it is the real user, it identifies them by matching their fingerprint and the image shown above.

The final confidential matrix containing the user's biological information and the saved data will be subjected to hamming code where additional bits will be added to detect and correct any errors that might occur to the stored biometric data as in the figure above.

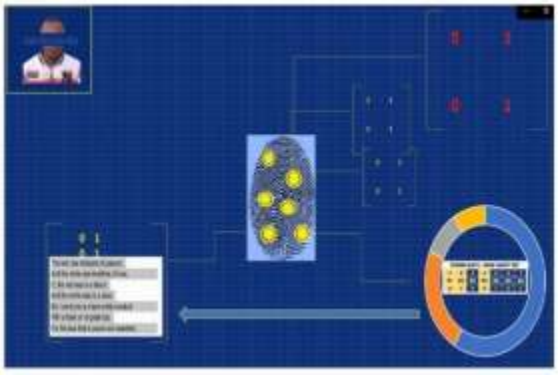


Figure 9: Final confidence matrix is hidden by text steganography (white spaces)

Once the parity bits are added, the message is stored using steganography. The presence of a white space is encoded as 1, otherwise, it is encoded as 0 as shown in above Figure.

CONCLUSION

Encryption, error detection, and steganography are possible techniques to secure biometric data. In this paper, the application of error detection through hamming code technique in the data is purposely recommended to secure the data presented in the biometric database. In addition to error detection, steganography encryption can also be used to increase the security of biometric data further.

ISSN:2789-3995
(Online)

Thus, using these techniques will not render the system vulnerable to the third party. Therefore, the combination of Hamming code and steganography will give more security and robustness will therefore decrease.

RECOMMENDATIONS

This study will enhance the security and usability of biometric technology. On the other hand, further research suggestions related to the current scope of the study can be considered for future work by using alternative methods to identify hash template functions within the proposed approach instead of the neural network technique to determine the accuracy of generating the confidential matrix. Additional work can also be performed by capturing enormous data of different fingerprint spectrums to showcase biometric modalities' behavior and transparency.

REFERENCES

Chi, L. (2016, April 11). Philippines elections hack 'leaks voter data. Retrieved from BBC.COM/NEWS: https://www.bbc.com/news/technology-36013713

Gao, J. (2007). Reed Solomon Code. Lectures on Wireless and Mobile Networks.

Hoffman, D. G. (1991). Coding Theory: The Essentials. New York: Marcel Dekker.

J. Menezes, P. C. (1996). Vanstone. Handbook of Applied Cryptography. CRC Press.

Jain, A. R. (2004). An introduction to biometric recognition. IEEE Transactions on Circuits and Systems for Video Technology.

John, B. (2022, May 26). A White Rose (n.d.). Retrieved from Poetry.com: https://www.poetry.com/poem/21979/a-white-rose.

Manual, B. A. (2008). Biometric Technology Application Manual Volume One: Biometric Basics. Compiled and Published by: National Biometric Security Project.

Moreira, J. C. (2006). Essentials of Error-Control Coding. West Sussex.: John Wiley & Sons.

Neil, Y. &. (2004). Fingerprint Classification: a Review. Pattern Analysis & Applications 7, 77-93.

Otti, C. (2015). Comparison of hand geometry and fingerprint-based identification. 3rd international conference and workshop Mechatronics in Practice and Education. Subotica, Serbia.

Rathgeb, C. &. (2011). A survey on biometric cryptosystems and cancelable biometrics. EURASIP Journal on Information Security.

Sam, K. (2021, October 14). High Court Declares Huduma Namba, Illegal. Retrieved from Business Daily Africa: https://www.businessdailyafrica.com/bd/news/high-court-declares-huduma-namba-illegal--3582926.

Schneier. (1999). The Uses and Abuses of Biometrics. Communications of the ACM, 136.

Tibor, K. (2015). Szerző, Biometrikus azonosítás. [Performance]. Óbudai Egyetem Bánki Donát Gépész- és biztonságtechnikai kar.

ISSN:2789-3995
(Online)

Tudor, J. (2001). Information Security Architecture. An Integrated Approach to Security in the Organization.

Uludag, U. ,. (2004). Biometric cryptosystems. Issues and Challenges, 948-968.

Wayman, J., Jain, A., Maltoni, D.,&Maio, DE E. (2005). Biometric Systems.