Automation, Control and Intelligent Systems
2021; 9(3): 89-96
http://www.sciencepublishinggroup.com/j/acis
doi: 10.11648/j.acis.20210903.12
ISSN: 2328-5583 (Print); ISSN: 2328-5591 (Online)

Bio-metric Encryption of Data Using Voice Recognition

Alhassan Jamilu Ibrahim, Usman Abubakar Jauro

Faculty of Science, Federal University Kashere, Gombe, Nigeria

Email address:
jamilualhassan@yahoo.com (A. J. Ibrahim), ausamn63@gmail.com (U. A. Jauro)

To cite this article:
Alhassan Jamilu Ibrahim, Usman Abubakar Jauro. Bio-metric Encryption of Data Using Voice Recognition. *Automation, Control and Intelligent Systems*. Vol. 9, No. 3, 2021, pp. 89-96. doi: 10.11648/j.acis.20210903.12

Received: June 27, 2021; **Accepted**: July 19, 2021; **Published**: August 19, 2021

Abstract: In symmetric cryptosystems, the protection of secret keys is based on the traditional user authentication and likewise the security of the cryptosystem depends on the secrecy of the secret keys. In the event of lost, theft or infection of these secrete keys; the security of the cryptosystems would be compromised hence exposing critical information. Biometrics has been commercially used to verify user's identity. Voice biometrics has been proven to be even more effective because it cannot be stolen in some cases like face, fingerprint or even iris biometrics. The research proves that a well-designed system will prompt an authentication question and on verification user must provide both the desired answer as well as desired matching threshold or the system ignores the user features. This research proposes a software-based architecture solution for Biometric Encryption of data using Voice Recognition that employed the Dynamic Time Warping (DTW) technique to solve the problem of speech biometric duration varying with non-linear expansion and contraction. The approach then used database to store the monolithically bind cryptographic key with the equivalent biometric hardened template of the user in such manner that identity of the key will stay hidden unless there is a successful biometric authentication by intended party. The research used the MIT mobile device speaker verification corpus (MDB) and A data set in quiet environment (QDB) for training and verifying session. Finally using the Equal Error Rate (EER) the research evaluated performance or rate at which False Acceptance Rate (FAR) and a False Rejection Rate (FRR) are equal. Therefore, according to the result it offers a better substitute method of user authentication than traditional pre-shared keys for benefit of protecting secret keys.

Keywords: Biometrics, Encryption, Decryption, Authentication, Device Pairing, Voice Recognition, Audio, Bioscrypts

1. Introduction

1.1. Biometrics

We can define biometric as the unique, measurable, biological characteristics or trait that recognize or verify the identity of a human being automatically. The science of biometric has been to statistically analyze these biological characteristics which are now typically used for security purposes these days [1]. For security installations, analyzed, five most common physical biometric patterns are fingerprint, hand, eye, face as well as voice.

It is clear that biometric has now become a commercial medium used in verifying user's identity which has become an alternative form of user authentication that is fast replacing traditional pre-shared keys or passwords. In this context, devices need access to keys before they communicate with each other. They do so by storing these keys in key rings [2]. Now when these keys are in store, possibilities arise for them to be tempered or got infected by devices to expose them and therefore compromise security. The science of biometric had provided solution to help solve this problem. This paper proposed a software solution to biometric encryption of data for storage using Voice Recognition.

Currently, there is steadily increase in authentication and identification requirements in both the online and offline modes. This has dropped the need on both the public and private sector stakeholders to be adequately informed on whom they are dealing with but which developing countries, in especially Africa, are still left behind [3]. Currently, models employed for security used in protecting information, verifying identity and authorizing access to premises or services depends on employing token, tied to and hence present individual, to either authenticate his/her identity or gain access grant to information, premises or services generally. The depiction here can be referred to an identity card (something you have), password or shared secret

(something you know), or a biometric (something you are). sampling all of the cases, you can agree that a third party gets hold of this token details and whose function is characterized by authorizing and sometimes proceed to grant transaction to commence or continue once details presented matches the one stored in the database. Although biometric is highly regarded as an ultimate form of identification and authentication, thus presenting the third and final component for proof of identity, it's still not water tight. Consequently, the concept is being employed for several security applications.

Privacy-related spans including protecting confidential data, though, proved not as powerful – biometrics are to satisfy this need. After a user provides their confidential data (financial, medical or otherwise) to a second party, and they will in turn normally specifies that it will merely use the confidential data for the designated purpose, and hence will proceed to protect the data from admission by unauthorized parties. The connection between persons who provides useful data and the successive party is mainly established on an ideal of trust. Over the time and due to technological advancements and geopolitical issues evolvement, trust model has become far less effective. The legal and illegal business of sharing and selling of personal information is nowadays a highly profitable business that is been widely practiced by many organizations. Consequently, with increasing threats of terrorism, various law enforcement agencies at different level of governments holds the right to demand access to more and more personal information. As internet's flexibility steadily expands, extensive character profiling can now be developed electronically about an individual, without his or her consent. The greatest threat lies where the person's life is adversely affected due to the errors that can easily arise.

Transaction details that are token-based for especially biometrics will be included in this report, in such a way that surprisingly includes the entire record of transaction histories for a person without their consent. Hence this will introduce the person's ability to access the database and make changes like correcting errors, thus presenting an ever-growing problem. Clearly, the unauthorized access of individuals personal information has grave risk which may include identity theft and its used in committing several offences on both the affiliated personal information as well as general crime.

Personal information is usually required by law enforcements in government with the intention of using it for own public safety or protection, and/or national security. In the other hand organizations [4] require personal information to improve business practices and customer service. Still despite all the existing devised models with protocols of protecting information along with privacy consistently leads to less security and more costly business practices. This issue can be reviewed.

It is necessary to protect public and nation's security, which has become mandatory function for any meaningful civilized society; whilst devising measures to enhance business practices that poster profitable gains as well as enhancing better customer service. The welfare or wellbeing both socially and economically is preserved with both of these functions in place. However, while considering decimating these functions, in reality, it is mandatory to also sight the liberty and freedom of choice for any prosperous free society. Over the years, the collection and processing of data or information using technological advances placed resources as vital to human health, well-being and freedom of people even more at risk. Clearly, the illegal or abuse of confidential data or personal information led to increasing threat, wasted resources, and generally lead, to some extent, the detriment of society [1]. Where individuals or society are paranoid or perpetually anxious about identity theft, misuses of their information, or unwarranted search and confiscations cannot efficiently function as expected.

The security model currently used optimized processes where the need for privacy and the protection of personal information as well as general security can be both be served Accordingly, this research proposed and important step in achieving that goal through a new positive-sum model for both protecting information and providing security, based on "Biometric Encryption using Voice".

1.2. Current Trend

The concept applied for biometrics are anticipated to contribute a new level of protection to request where a person endeavoring access have to clarify who he or she truly is by giving a biometric to the system. These arrangements could additionally have the ease, from the user's outlook, of not needing the user to recall a password. There is need for creating awareness and attention in the use of biometrics [5]. Border Control: Perhaps the most visible (and controversial) use of biometrics is read-through area in the transportation sector in advanced societies and especially emerging countries of Africa. Identification necessities at airports and border crossings could nowadays involve the collection and processing of travelers' fingerprints, facial pictures, and iris patterns. Increasingly, mechanism readable travel documents such as passports, driver's licenses and supplementary individuality or travelers' cards could additionally include biometric data or images. Regular travelers who apply for and bypass comprehensive background checks could use their biometrics for speedy method across customs and immigration. In crime and fraud prevention, detection, and forensics: The use of biometrics such as fingerprints by law enforcement has taken place for years, but nowadays that fingerprints can be digitized, stored, retrieved and matched instantaneously and easily manipulated, countless new uses have appeared, such as for populating detection of catalogs and grasping out confidential sector background checks [6]. In some sectors of advanced countries, cashing cheque will need biometric imprint to be allocated on the front side.

With time technological advancement has described a little new "revolutionary" biometric knowledge that promises to resolve crime and usually make the globe a better place to live. Attendance Recording: Employees and students are being needed, in generating numbers, to present a biometric

(such as a finger or face) in order to "check in" to premises, far like a punch timepiece, or to claim a little claim such as a luncheon meal or to check out a library book. Payment Systems: We are witnessing increasing uses of biometrics by the confidential sector for enhanced ease services, such as "pay 'n' go" arrangements that permit registered clients to pay for groceries or gasoline using merely their finger at periods, a large convenience. Admission Control: One of the most extensive uses of biometrics has been for physical and logical admission to safeguard distances or resources (e.g. to a database of health records, or accessing a laptop). In such conditions, biometrics can enhance protection by helping to safeguard that admission to sensitive resources is severely restricted to authorized individuals [5].

1.3. Biometric Premier

"Biometrics" denotes automatic systems that use quantifiable, physiological, physical characteristics or behavioral qualities to comprehend the individuality, or verify/authenticate the personality of an individual. Some of the illustrations of biometric features that have been exploited for automated recognition contain hand or finger geometry, retina, fingerprints, iris, face, voice, signature, and keystroke dynamics. These systems are conventional on the following steps: a biometric data samples are taken from an individual's biometric captured data. The presented physical characteristic could be an image or other biodata captured often where data are extracted from that sample [7]. This extracted user information comprise of what is referred to as biometric template. The biometric data, whichever the picture or the template or both, are next stored on a database or a distributed nature, such as intelligent cards. These preparatory periods equally contain the procedure of enrolment. It is worthy to note that the person whose data are stored is referred to as the enrollee.

The definite aim of the biometric system is purely attained at a larger level. If a person presents herself/himself to the system, the system will ask to present his/her biometric features. The system will match the image presented sample (or the template removed from it) alongside the biometric data of the enrollee [9]. If the match succeeds, the person is then verified and the system will "accept" his/her. If the match does not thrive, he or she is not recognized and he/she will be "rejected."

1.4. Problems Associated with Biometric Identification

Let us site example from the hi-tech film 'Minority Report' starring Tom Cruise, where in the movie individuals are automatically and instantly identified via remote scan of their irises. To beat the system and escape detection, characters must literally change their eyeballs when defying the system. However, such scenario isn't likely to happen many real-time environments because, for various reasons, biometric technologies are not well matched for large- scale one-to-many real-time identification resolves. It is important to keep in mind that the collecting biometric samples or data and their designation into biometric templates for desired matching is oftentimes subjected to variation. For clarification, biometrics is "fuzzy" – thus it is impossible for two samples to appear perfectly identical. For example, biometrics authentication using Facial recognition may be prone to variability due to different lighting as well as other variant conditions, angle, subject movement, and so forth [7]. For this reason, it can be remembered where passport photograph shots are generally not allowed to smile. Similarly, numerous reasons affect the ability to obtain reliable fingerprint samples. Perhaps, one of the most reliable and almost perfect biometric authentications tend to be irises. For this reason, collective biometric samples can be at some variance with stored reference samples, making comparison, matching and identification an inaccurate procedure. In other words, there is no biometric system that is proven to be 100 per cent accuracy. When the biometric system cannot perform a proper match and (incorrectly) rejects a genuine user, this is called a false reject, and such genuine user must have to typically resubmit another or other biometric samples or information for further evaluation by the system. In the identification scenarios [8], the biometric effectively serves as an index or key to the database involved much like systems where login usernames aid to recognize registered users of a computer network [11].

Reputable organizational structures, systems, secured facilities and their likes are in constant demand for improved security systems that mostly uses scan and/or biometric authentications in order to increase the layers of security. Humans in general are always in search of simpler machines, in order word, technology that will further simply the ever-growing human tasks and also improve efficiency of their outputs. This had push man to inventions that will employ biometrics, like voice authentication, in order to improve command between man and machine. Despite efforts from researchers, these systems had remained rear in especially the third world countries.

This research took a dimension that will suit even the diverse African dialects and the emerging development and their environments at large. As mentioned, availability of such technologies is practically rear in especially these third world countries while the research will device a cheaper system that can be affordable to more lower income parties.

2. Literature Review

2.1. Biometric (Voice) Encryption

In Biometric Encryption, you can use the biometric to encrypt a PIN, a password, or an alpha-numeric string, for numerous applications to gain access to computers, bank machines, to enter buildings, etc. The PINs can be 100 s of digits in length; the length doesn't matter for you don't need to recall it. And most basically, all one has to store in a database is the biometrically encrypted PIN or password, not the biometric template. Biometric modalities can be presented in numerous forms (e.g. fingerprints, iris, voice

etc.) and each distinctive feature can be used to produce cryptographic keys. Unlike physiological biometrics, behavioral biometrics changes with actions that are been executed, therefore attractive for key generation by allowing users create different keys.

It was Monrose *et. al*. who suggested the use of the first applied system which implements behavioral biometrics to generate key [9]. They argued that their work produces keys as robust as passwords by generating latencies which increases the entropy of standard passwords. At a point their proposed method will increase the workload of an adversary by a multiplicative factor of 215. The most essential feature of the edifice is how they identify distinctive and useful keystroke for a specific user. They assign global threshold that extracts 0 or 1 bit for each feature of user that are consistently below or above that threshold. On verification user must provide matching threshold or the system ignores the user features.

RBTs are in a way similar to *Monrose et. al.* [10] but with more other advantages. Although they implemented quantization for error correction, they can still achieve higher entropy rates and lower false accept rates because they partition the range of each feature into more than two segments. Another thing is their idea of distinguishing feature they select to allocate to each user is even more flexible. When we make comparison, it is clear that Monrose's approach ignores important features that can reliably be repeated, but then the mean of these features falls on the global threshold. Lastly their work was computationally more efficient than number theoretic primitives because their approach was based on block ciphers and hash functions.

Other works are presented in the area of "Fuzzy Cryptography", an idea presented by Juels and Wattenberg [11] to support noise-tolerant. Then there is Fuzzy Vault, case in point of Secure Sketch implemented to build Fuzzy Extractor. They projected an approach which assesses BKGs by investigating the use of weak forgeries to analyze BKGs for better security. Their work also explored "generative algorithms" which creates forgeries using partial knowledge on user's biometric.

Lastly, they proposed algorithm that counts key space to find a user's key by probability and defined a set of necessary security requirements. This paper will adopt a different methodology that will also deliver necessary security requirements, with addition to a security that will almost be impossible to break.

2.2. Voice Biometrics

Using computer security terms, we can define reference monitor as an abstract machine that oversee all access to entities by subjects within a computer system. Tremendous Work was made in the areas of implementing speech-based reference monitors which controls the access interface between humans and computer resources. In voice biometrics, speaker authentication technologies are implemented as typical reference monitors and they include automatic Speaker Verification (SV) and Verbal Information Verification (VIV) [12]. The term SV is used to refer to verifying if an unknown speaker is really the person, he/she claims to be while VIV verifies spoken utterances compared to the stored information on a given personal data profile. Performance varies in these approaches, using various scenarios but all these replaced traditional passwords for authentication and verification of users. They are particularly convenient using applications where speech is the natural human device interface.

Cryptography using voice biometric can be implemented to encode message or data [13], making it unintelligible to anyone other than the intended recipients (Confidentiality). Also, it can be used to encode the message or data so that any attempt to modify by an unintended party can be detected (Authenticity). Even when an attack-resistant reference monitor cannot be implemented, attacker who intercepts the encoded data cannot make any sense of it needless to modify it without detection. The ability to decrypt an encoded data to an intelligible form or encode data that will look authentic requires a secrete value called the key which must be fed to the encryption algorithm. For the cryptographic algorithm to properly functions, this secrete key must possess certain properties.

1. Key must be unpredictable by unintended parties.
2. Key must be reproducible by intended parties.

2.3. Vioce-Generated Cryptographic Keys

Several researchers advocate research in areas of generating cryptographic keys using voice input. With particular interest in unpredictable keys which are reproduced temporarily on the same device from the same user's voice. This approach is harder to implement than the speech-based reference monitor because solutions to problems can be used to construct the reference monitor directly. Here is an approach where the cryptographic key derived from the voice input/signal will be received by the reference monitor and then compares to what the key was actually supposed to be (similar to what password-based login program does).

When you use automatic speech recognition to verify spoken password and then use it (password) as key, you might have achieved the goal of unpredictable using possibly the most accepted approach to developing a repeatable key from a spoken passcode. Modern ASR is employed, a tool that recognizes vocabulary of almost 104 words under the best conditions [12]. It is easy of cause to search using automated attack when a key is drawn from such small space. As experienced with PINs established, when it is required by the user to create password by adding several words that taxes the memory of the user it will produce marginal improvement in entropy. It is necessary to draw entropy from how the user speaks a password which is in contrast on how unpredictable is achieved. Now, let's draw a solution to this problem.

A user is prompted for password (or passphrase) and then he/she will utter to the device. The device will use the voice input to generate a sequence of equivalent bits (key). When the user utter same password several times it should be able

to generate the same key for verification. At the same time if an attacker was able to dissect the system, he/she will be landed with horrible work of recovering the key. Ideally, even when the attacker knows the password the key should be able to resist discovery. It is important to highlight here that models used for voice verification commonly used in speech-based reference monitors like voice dependent model or the text dependent models are likely to leak important information to an attacker that successfully dissects the device that stores the model.

For secrete passwords, a text dependent model can vulnerably leak information of these passwords while the voice dependent model usually leaks information of voice features that is pertinent to the user. The device must use speaker's voice features to generate a speaker dependent cryptographic key without referring to either the text or speaker dependent model. Based on my research, false negatives are appropriate for key reproducibility while false positives aren't the most significant measure for the security of the system (unpredictable) [12]. It is clear that false positives don't capture the full problem postured to the attacker who captures the device, rather, more apprehensive with security of the system. This paper will present a similar but more robust measure and its implementation in experimental evaluation**.**

3. Approach

The researchers will be employing the Dynamic Time Warping (DTW) technique to solve the problem of speech biometric duration varying with non-linear expansion and contraction. The technique sets a nonlinear mapping between two signals by minimizing the distance between them. To setup a warping function for incoming inputs we need template for keying signal while this information must be stored to a secured template like smart card to avoid leaking information. Information stored may include time domain features such as Energy and Zero Crossing Rate [10], which cannot be inverted to reconstruct the key. In this case robustness is reduced, frequency domain features that can be inverted to original signal. The research also proposes a stored template based on frequency-domain features that is almost or impossible to be inverted to its original form used for DTW based user authentication system. To upset the originality of the template, the hardening algorithm is used to remove some frequency-domain features from the template and then transforming the rest of the features to a time-domain template. This template is used as a keying signal in the dynamic time warping (DTW) [15]. To create a hardened or stored template, the Discrete Fourier Transform and the inverse DFT are used. Again, we are faced with the other problem which is the correlation among features. Authors reported that "an iris code usually has a run length of 8 consecutive '1's or '0's." which means that binary do repeat results [10]. In speech, the length of repetition depends on the number of phonemes in a pass-phrase and the idiosyncrasy of users as they produce their spoken password. The researchers employed a mapping algorithm to address this issue by implementing multi-thresholds, determined from pseudo-random bits. The end result of the algorithm will be binary string generated in such a form that an adversary cannot predict. In a nutshell, the approach will focus on how to transform speech reliably, securely, and randomly into binary string. The two entities, DTW to make the scheme more liable and hardened template to maintain security, will see to the progress of the approach. Lastly, binary string will be generated unpredictably to maximize the entropy of the template through the help of the multi-thresholds scheme.

3.1. Speech Processing

The function S_a(t) usually denotes a speech signal where t is the time. The first step is analogue to digital conversion where the analogue is sampled at 8 *KHz*, reconstructing signals less than 4 *KHz* according to sampling theorem. To strip the higher frequencies from this signal we implement the low-pass digital filter with cut-off at 4 *KHz*. Let P be the sampling period hence our digital signal can be deduced as $S(n) = S_a(n.P), n = 0, 1, 2 \ldots . N-1$. To raise the Signal to Noise Ratio, the signal is passed to 1st order digital filter $H(z) = 1 - \alpha z^{-1}$, α ranges between 0.9 to 1 [12]. For the next step, the signal is framed into short analysis interval so that each frame is multiplied by a window function which will reduce any sudden changes at the beginning and at the end of the frame. Length of each frame is usually 30 *msec* and must overlap properly, usually 10 *msec* overlap. The last step will be to extract these features from the signal.

As widely practiced, basic voice features are in the Discrete Fourier Transform (DFT). The Discrete Fourier Transform of N points signal $X(n)\ for\ k = 0, 1 \ldots . N-1$ can be defined as:

$$X(k) = \sum_{n=0}^{n-1} X(n) exp \frac{-j2\pi nk}{N} \quad (1)$$

The inverse transform for $n = 0, 1, 2, \ldots . . N-1$ can be defined as

$$X(n) = \sum_{n=0}^{n-1} X(k) exp \frac{-j2\pi nk}{N} \quad (2)$$

According to the real function property, if $x(n)$ is real and $x(n)$ and $X(k)$ are transform pairs, then

$$X(-k) = X(N-k) \quad (3)$$

This symmetric property, equation (3), can be exploited to decrease the computation required to transform a real sequence. To derive DFT, there is no need to compute $X\ for \frac{N}{2} < K < N$, since these values can be found from the first half of X. Cepstral coefficients or cepstrum, are the most efficient features used in identifying a speaker. Cepstrum physically represents the movement of articulators (the teeth, alveolar ride, hard palate, and velum) of speakers. Its use is popular because of low correlation. Hence, it is appropriate to apply it for a cryptographic purpose. Cepstrum can be defined as the Inverse Fourier Transform of log-energy of Fourier Transform [13] of a signal $S(n)$. By definition $S(n)$ where $C(v) = F^{-1}\{\log|F\{S(n)\}\}|2$ and $S(n)$ where F^{-1} and F denote Fourier and Inverse Fourier Transform.

3.2. Biometric Key Generation

As practiced by previous designs, this overview will have training and verifying sessions, the training session is where the biometric key regeneration takes place where users will repeat their training passphrase $i + 1$ times for the system to successfully derive the cepstrum and DFT features. This process involves digital signal processing, using 240 samples per frame that are shifted every 80 samples for the sampling rate of 8 *KHz* [14]. A total of 12 cepstrum and 121 DFT features are contained in each frame. The stored DFT features of m frames is used to initialized the system from one of the training utterances, referred to as the DTW template, carryout the DTW on the rest of the training utterances. For the mapping process the cepstrum features derived from the warped signals are used by mapping each utterance to a binary string of length m called a feature descriptor. Finally, each one of the feature descriptors are employed to designate a unique feature the user can consistently generate.

Let RP and DT be equals to the bits number derived from random pass-phrase and a DTW template respectively then we have DT produced by the system to be less than or equal to the fixed threshold (RP). To define the distinguishing features, we use the RP threshold derived from the full template experimented by keying signal to the template. The RP threshold is the number of bits corresponding average distinguishing descriptor of the random passphrase.

The distinguishing descriptor D should exceed the experimental thresholds which is greater than the time-domain distinguishing descriptor TD from using a time-domain feature template. However, the distinguishing descriptor D cannot exceed FD, the frequency-domain distinguishing descriptor from frequency-domain full template. Therefore, thresholds lie between TD and FD. Let the suitable threshold be equals to T, so that if $DT \leq RP$ and $D > T$ is likely not to leak information to an attacker using random pass-phrase attack. To achieve this the template is hardened and DT is made greater than RP so that the system finds suitable length of distinguishing feature from the biometric by rejecting D less than or equal to T. the conditions will be re-started until they are met, after DTW is made keying signal of the training passphrase following each step in hardening the template. Lastly, IDFT from the hardened DTW template is stored as the hardened template and $2n - 1$ of descriptor, where n will be selected based on feature variation to from binary string S, thus $n = 3, 4 \ldots\ldots$

Let $E(k)$ be the random key generated and then properly encoded after the hardened template is set. Using BCH code, we can deduce that the encoding code $E(k)$ has tolerated error within Hamming distance (H), maximum bit difference and descriptor of a legitimate user. Next, we use an XOR operation to hide the distinguishing descriptor S and the encoding code $E(k)$ then stored as lock data (L). Now to unlock L will require user with feature descriptor S' that is exactly similar to that of the Hamming distance ($|S - S'| \leq H$), hence can move to decode the key. This is can be termed as fuzzy commitment [11].

3.3. Multi-Threshold Generation

Let's maximize the entropy, the distinguishable entities to a biometric system, of the biometric template by selecting sets of thresholds [14]. Therefore, a set of thresholds used in the mapping process is supposed to produce binary string which will appear random in cryptographic concept. Using algorithm in [7], we generate pseudo-random bits $p \in \{0, 1\}^m$ and then successfully deleted after a set of biometric might have successfully been selected. The mapping algorithm matches the feature greater than the threshold by '1' and '0' otherwise, therefore we set a threshold lower than the mean of features when corresponding pseudo-random bit is '1' and greater than mean otherwise. Let $\mu_j(i)$ and $\sigma_j(i)$ be the mean standard deviation of the linear combination of the cepstrum features i^{th} of frame over l training utterances which enable us to generate multi-thresholds for user j. Below are steps of how the algorithm executes.

1. Generate pseudo-random bits $p \in \{0, 1\}^m$ using algorithm [7].
2. Set the multi-thresholds $T_j(i) = \mu_j(i) + \left(-1^{p(i)}\right)k\sigma_j(i); k > 0$
3. Securely delete pseudo-random bits.

3.4. Biometric Mapping to Binary String

The cepstrum features are mapped to corresponding binary string using the algorithm below, as well as to define D, the distinguishing descriptor for user j having l training utterances.

1. Perform DTW to the training utterances.
2. For each frame of k^{th} training utterances, let $f_{j,k}(i)$ represented the cepstrum feature, where $i = 1, 2, \ldots\ldots m$ is the number of frames. Compute $f'_{j,k}(i)$ from the linear combination of $f_{j,k}(i)$.
3. Generate multi-thresholds $T_j(i), i = 1, \ldots\ldots m$ using the algorithm in Section 3.1.
4. Compute the i^{th} feature, $\varphi_{j,k}(i) = f'_{j,k}(i) - T_j(i)$.
5. Binarize $\varphi_{j,k}(i)$ to the feature descriptor, $b_{j,k}(i)$ by testing whether $\varphi_{j,k}(i)$ is positive or negative. Map to '1' if it is positive and '0' otherwise.
6. For the training utterances, determine XORing of, $b_{j,k}(i)$ for $k = 1, \ldots\ldots l$. If the XORing of $b_{j,k}(i)$, is zero, the i^{th} feature will be distinguishing feature and set $B_j(i) = b_{j,k}(i)$, otherwise $B_j(i) = \tau$.
7. Determine D, the number of bits that $B_j(i) \neq \tau$. If D is less than or equal to T, reject the biometric.

3.5. Hardening Template

Using DTW technique we can store a hardened template so as to set time alignment to input signal which should not be transformed to original template of m frames of 121 features each. It could be easier to choose optimal features that should yield $DT \leq RP$ [15] by enumerating over m frames of this original template, but the computational time isn't feasible. We are left with the option of employing the optimal search

algorithm. These researchers will be employing the Sequential Backward Search (SBS), a top-down search procedure that eliminate one feature for full set of features per step until condition is met. We can employ it to compute and eliminate, on each step, the DFT feature that maximizes DT until DT is less than or equal to RP. Below is the breakdown or description of the algorithm for hardening a template in steps:

1. Initialize by setting one of the training utterances as a DTW template, a set of DFT features.
2. Remove one of the features from the DTW template that minimizes DT by performing the algorithm in Section 3.2.
3. While $DT > RP$ go to step 2.
4. Terminate, the IDFT of the remaining features is stored as the hardened template.
5. Define $2n - 1$ the least variation of the distinguishing features, where $n \geq 3$ to form binary string S.
6. Set the lock data, $\mathrm{C} = E(k) \oplus S$, where $E(k)$ is the encoded key and $\oplus$ denotes XOR operation.
7. Securely delete a set the training utterances.
8. Store C, T_j and the hardened template in the database.

3.6. Biometric Key Retrieval

This is the process of verification where the user makes a request of the template from the system's database containing the hardening template, multi thresholds, as well as the lock data. Signals from the DTW of user's pass-phrase are executed to generate the feature descriptor and XORed with lock data. We now move to the decoding process where the key can be reconstructed correctly if error is within Hamming distance. We check the hash function to verify if key is similar to key generated in the training phase [15], where it was stored as $h(k)$ initially. The system now performs verification computation to find out if

$h(k) = h(k')$. If $h(k) = h(k')$, the key, k', is correct.

4. Experiments and Results

Using the Equal Error Rate (EER) we evaluate performance or rate at which a False Rejection Rate (FRR) and False Acceptance Rate (FAR) are equal. The percentage of time the system accepts the wrong speaker or unauthorized system user is FAR. Consequently, the percentage of time the system rejects an authorized speaker is FRR. Following thus we implement two databases for this experiment: The MIT mobile device speaker verification corpus (MDB) and A data set in quiet environment (QDB). The public database obtainable by MIT is MDB, while QDB is the database composed over the period of a month.

For MDB, four times repeated pass-phrase is employed as the pass-phrase to be used for training which is also used for verification pass-phrase. Similarly, for QDB, for each speaker in the experiment we use five pass-phrases which will total 5*6 = 30 distinctive pass-phrases. To evaluate the imposter trial, five recordings same pass-phrases from the speakers are used. For each pass-phrase they will have a total of 25 recordings (5*5). Next, the random pass-phrase trial is evaluated by selecting 25 other pass-phrases not corresponding to the verification pass-phrase and from other speakers at random. We then set the length of binary string to 127 and 255 bits for MDB and QDB respectively and use BCH code to enable us set the code word to 127 and 255 bits.

The table below compares the performance of the full, the time-domain, and the one-way function template with the equivalent hardened template.

Table 1. *Hardened template Vs Full, Full-time domain, and One-Way function templates.*

Database	Template	EER (%)	
		Random	Imposter
MDB	Full	3.62	12.75
	Hardened	4.43	13.14
	Time-domain	6.22	15.59
	One-way	8.24	21.44
QDB	Full	2.87	12.87
	Hardened	3.80	13.80
	Time-domain	5.60	15.20
	One-way	7.13	20.27

Table 1 shows that using MDB and QDB to recognize the performance of the hardened, the full, the time-domain, and the one-way function template is a success. The results clearly indicate that the EER for the employed scheme outperforms the time domain as well as the one-way function templates.

5. Conclusion

The paper presented a solution to the various problems associated with the safekeeping of cryptographic keys by using a biometric (voice) method of secured authentication to generate and manage these keys. Also, the approach tackled the problems associated with feature correlation, increased randomness of the key and also addressed the challenges associated in using the DTW in cryptosystems. The paper specifically proposed the hardened template to preserve that fact that when used to create wrapping function, they cannot be transformed into original template or reduced in performance. Generally, the security and convenience of using this biometric method of encryption will surely promote widespread attention for cryptographic systems.

References

[1] Cavoukian, Ann, and Alex Stoianov. Biometric encryption: A positive-sum technology that achieves strong authentication, security and privacy. Information and Privacy Commissioner, Ontario, 2007.

[2] Soutar, Colin, Danny Roberge, Alex Stoianov, Rene Gilroy, and BVK Vijaya Kumar. "Biometric Encryption™."

[3] J. R. Deller, Jr., J. H. L. Hansen, and J. G. Proakis. Discrete-Time Processing of Speech Signals. Macmil- land Pub. Co., New York, 1993.

[4] Chandra, Sayani, et al. "Generate an Encryption Key by using Biometric Cryptosystems to secure transferring of Data over a Network."

[5] S. Furui. Digital Speech Processing, Synthesis and Recognition. Marcel Dekker, Inc., New York, 2001.

[6] F. Hao, R. Anderson, and J. Daugman. Combining cryptography with biometrics effectively. IEEE Transactions on Computer, 55 (9): 1081-1088, September 2006.

[7] A. K. Jain, K. Nandakumar, and A. Nagar. Bio- metric template security. EURASIP Journal on Ad- vances in Signal Processing Special Issue on Biomet- rics, January 2008.

[8] A. Juels and M. Sudan. A fuzzy commitment scheme. In Proceeding of the 6th ACM Conference on Computer and Communication Security, pages 28-36, November, 1999.

[9] T. Kinnunen. Spectral Features for Automatic Text Independent Speaker Recognition. PhD thesis, Department of Computer Science, University of Joensuu, Findland December 2003.

[10] Monrose, Fabian, et al. "Cryptographic key generation from voice." Security and Privacy, 2001. S&P 2001. Proceedings. 2001 IEEE Symposium on. IEEE, 2001.

[11] Venkatachalam, S. P., P. Muthu Kannan, and V. Palanisamy. "Combining cryptography with biometrics for enhanced security." Control, Automation, Communication and Energy Conservation, 2009. INCACEC 2009. 2009 International Conference on. IEEE, 2009.

[12] F. Monrose, M. K. Reiter, Q. Li, D. Lopresti, and C. Shih. Towards speech-generated cryptographic keys on resource constrained devices (extended abstract). In Proceedings of the 11th USENIX Security Symposium, August 2002.

[13] Inthavisas, K., and N. Sungprasert. "Cryptographic Key Regeneration from Speech." Proceedings of the World Congress on Engineering. Vol. 2. 2013.

[14] Carrara, Brent, and Carlisle Adams. "You are the key: generating cryptographic keys from voice biometrics." Privacy Security and Trust (PST), 2010 Eighth Annual International Conference on. IEEE, 2010.

[15] R. H. Woo, A. Park, and T. J. Hazen. The MIT mobile device speaker verification corpus: data col- lection and preliminary experiments. In Proceedings of Odssey, The Speaker and Language Recognition Workshop, San Juan, Puerto Rico, June 2006.