Link Layer Type-Tagging in Fixed-Phrase Speaker Verification Systems (May 2005)

Mr W. Oyomno[1], Dr H.S. Venter[2]
Information and Computer Security Architectures (ICSA) Research Group
[1]were@tuks.co.za, [2]hventer@cs.up.ac.za

Abstract—This paper presents a design principle that reduces the ability to launch replay attacks on fixed-phrase speaker verification (FPSV) systems. The FPSV system design is based on the security mechanisms implemented at the network link layer. Speaker verification is a biometric technique, used to confirm the identity of a person, based on his or her voice characteristics. The term 'fixed phrase' indicates that the enrollment voice utterance and a test voice utterance match and, therefore, refer to exactly the same phrase. The presented design principle extends existing FPSV systems by introducing an anti-replay-attack component. This extension of the FPSV system does not have a significant effect on either processing power or storage of the system. In particular, type-tagging of a voice utterance with an anti-replay-attack tag is discussed. The tag achieves two important goals. Firstly, it specifically prevents an impostor from lodging a replay attack against the sensor inputs of speaker verification systems. Secondly, it confirms that the authenticating user is indeed a living person, i.e that the user satisfies the liveness property. The liveness property seeks to establish that the biometric sample presented for verification, is from a living person present at that instance it is presented. Liveness detection is based on the recognition of physiological or behavioral information as signs of life. Our focus is not on whether the design guarantees absolute security of the FPSV system, but rather on proposing a new alternative to improving the robustness of the FPSV system at relatively low cost.

Keywords—Fixed-phrase speaker verification, replay attack, speaker verification, liveness property,

I. INTRODUCTION

A Replay attack on a speaker verification system describes the misuse of a legitimate voice utterance or parts of the utterance to impersonate the legitimate user [1]. This voice utterance is taken from the context of a particular protocol run by an imposter. A protocol run refers to the successful verification of a claimed identity by an authentication system. There is currently no implementation of the fixed-phrase speaker verification (FPSV) system that specifically addresses this replay attack vulnerability of the FPSV system. The recommended solution is a prompted-phrase speaker verification (PPSV) system. A PPSV, as the name suggests, is a text-independent verification scheme, which necessitates the user at registration to utter a large amount of training phrases. The large size of training data results in a more sophisticated, as well as a longer registration procedure. A long registration procedure is deemed necessary to enhance unpredictability of test utterance requests. This unpredictability means the claimant may be prompted to randomly utter different phrases at subsequent verification requests. Despite PPSV being immune to the traditional replay attacks, it is computationally demanding on both storage space and processing power. Replay attacks on speaker verification systems are prevented by affirming that the test utterance being captured is a genuine utterance from the authorized live person who is present at the time of capture. The liveness detection aims to prevent the presentation of a recorded voice utterance or a dismembered body part for authentication by an imposter. This liveness problem justifies the need for a speaker verification system based on a fixed-phrase mechanism, which offers considerable immunity to replay attacks while using less computing resources [2].

This paper presents a design improvement on the FPSV system that not only reduces the ability to launch replay attacks against the system, but to also keep the computational space and processing power to a minimum. The general idea is borrowed from two fields, the first is the network link layer security architecture; the checksum mechanism [3]. The second field is signal processing. In particular the concepts of audio watermarking and audio fingerprinting have being adapted giving rise to a relatively new idea; "speech fingerprinting". Audio watermarking describes the use of special signals embedded into digital audio. These signals are then extracted by detection mechanisms and decoded. This mechanism relies on the limitations of human auditory systems within a particular audio range [4] [5] [6]. Audio fingerprinting on the other hand addresses the problem by finding and locking on features of audio that are invariant to time, frequency distortion and audio modifications, both unintentional and intentional. Audio fingerprinting makes great use of Two-dimensional features, called joint acoustic and modulation frequency [7][8][9][10]. It is this speechfingerprint that has being summarily referred to as the anti-replay attack tag. To address the susceptibility to replay attacks, the presented design has taken into consideration the liveness property.

In the next section the background upon which this paper is based is presented. The proposed design model is then discussed in detail in Section III. Section IV critically analyses the model and addresses issues relating to its use. The overall conclusion is presented in Section V.

II. BACKGROUND

A biometric is an attribute that really characterizes the given person; it represents the oldest form of recognition by humans and animals. The biometric authenticating system is a tool that verifies an individual's authenticity based on his or her distinguishing physiological and/or behavioral characteristics. Because most physiological or behavioral characteristics are distinctive from one individual to another. Biometric authentications are more reliable than either knowledge-based or possession–based authentication techniques in differentiating between authorized persons and impostors. It is therefore paramount that the used biometric characteristic is contrasted to those of the wider population. An ideal biometric should therefore possess the following vital characteristics [11]:

- It should be universal, such that each person possesses the characteristic.
- It should be unique in that no two persons should share the characteristic.
- It should be collectable, because the characteristic is readily presentable to a sensor and is easily quantifiable.
- It should be permanent, i.e. the biometric characteristic should not change or be easily alterable.

There are numerous forms of biometrics to choose from and naturally some are more reliable than others depending on the intended use. Some forms of biometrics include hand geometry, fingerprints, iris scans, DNA, typing patterns, signature geometry, body odor and voice biometrics [12]. These are illustrated in Fig1, which shows the various types of biometrics and their relationship with one another.

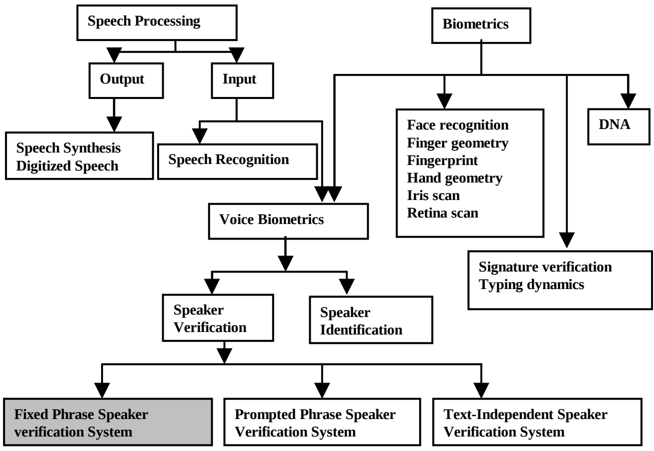

Fig1. Relationship between voice biometrics and other biometrics

The focus of this paper is on voice biometrics - in particular speaker verification. Speaker verification not only satisfies the requirements of an ideal biometric, it is also user-friendly and practical to use [12]. Speaker verification is a subset of the voice biometrics class.

In the next section basic issues related to speaker verification are discussed.

A. Speaker Verification

A traditional speaker verification system typically comprises of two very distinctive stages: enrollment and verification. At the enrollment stage, a user's training utterance is recorded and processed into a template stored in the database. During verification, one presents a test utterance that is verified against the enrollment template. The speaker verification system, thus, has three components facilitating these two stages: the front-end processing, the training of the speaker model and pattern-matching, as shown in Fig2.

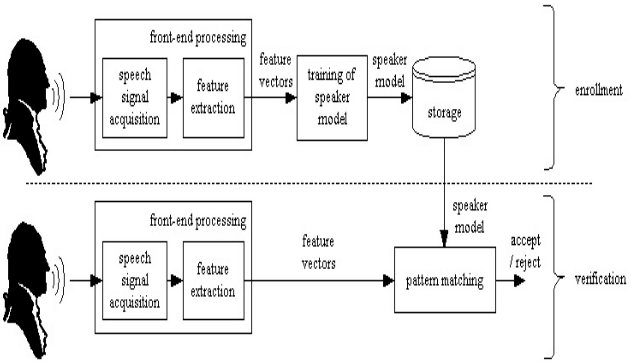

Fig2. Enrollment and verification stages in a typical speaker verification system [13].

The front-end processing is further divided into speech signal acquisition and feature extraction [13]. A typical FPSV process follows three stages:

1. The distinctive voice trait (feature) is extracted from the presented test utterance.
2. This trait is compared with a registered template using a pattern-matching algorithm.
3. Based on the similarity between the test trait and the template the instance is either rejected or accepted.

Nickel [14] attributes the success of a speaker verification system to its discriminative power and the particular pattern-matching algorithm used. These pattern-matching algorithms are optimally designed. Thus, there is not much room for performance gain on speaker verification systems' pattern-matching algorithms. Therefore, the performance of the system could be improved in other areas such as in the hardware of the system [14]. A speaker verification system quantifies unique signal characteristics of a vocalized phrase. The vocalized phrase is an alterable biometric. It is alterable because the same speaker can vocalize different phrases. An alterable biometric is similar to a password: if a password is compromised, one simply changes the password. Likewise, if the password is compromised, one simply needs to change the phrase. However, with an inalterable biometric e.g. a fingerprint, if the fingerprint is compromised one can not simply change the fingerprint. Thus with an inalterable biometric the options are limited, short of abandoning the use of the biometric system. The alterability of voice biometrics as well as the potential to detect duress, intimidation, compulsion or coercion through the analysis of stress patterns in the uttered test voiceprint, justifies choosing it as a preferred biometric [15], [16].

B. Type-tagging

Link-layer integrity of network communication is maintained using a checksum mechanism of the network communication [3]. This mechanism detects bit errors during transmission. A checksum is preferred in network communication, because it requires very little packet overhead and it is easy to implement in software. Other concepts we have borrowed from are audio watermarking and audio fingerprinting – to maintain integrity. We have borrowed both these concepts to design an integrity check in a biometric context. In a non-biometric context, replay attacks are deterred by type-tagging messages with unique cryptographic functions. These cryptographic functions, for instance timestamps, bind the message and components of the message to the correct context. The tag contains adequate information to clarify to which state of a protocol run the message belongs to. This tag ensures that the message integrity is protected; otherwise it does not increase the security of the protocol. Thus, tagging as technique improves a protocol's design robustness more than it guarantees the security of the protocol [17]. Tags are categorized on the basis of the primitive type of information they contain [2], [17]. The primitive types identified are:

- Protocol identifier, e.g. the message is authenticated because its protocol is authentic.
- Transmission step identifier, e.g. the message is genuine because information about its transmission have being verified.
- Message subcomponent identifier, e.g. the message is proved authentic because it contains information that verifies its other components as legitimate.
- Primitive type of data items, e.g. we can trust the message as it contains actual data necessary for a particular protocol run.
- Protocol run identifier, e.g. the precise use and details of the protocol runs are presented for verification.

Except for the protocol run identifier, the tag primitives as listed above are static in nature. Their static nature hampers their ability to fully accommodate the dynamic nature of vocal characteristics in a speaker verification system. This dynamic nature of voice characteristics is attributed to the fact that, unlike passwords authentication, it is difficult to conceal one's voice. The inability to conceal this seemingly sensitive information has resulted in voice being considered an open secret in a speaker verification context.

C. Open secrets

Most biometric authentication systems treat biometrics as secret information similar to passwords. Such systems, however, only grant access if the biometric sample provided matches with the stored sample. This rigidity in the design of biometric systems often implies that they are unable to deal with situations where the biometric feature is compromised. A more flexible approach is to consider biometrics as sensitive data requiring adequate protection, but not as secrets. This advocates for robustness in the biometric system design, such that it anticipates compromised biometrics and reacts accordingly. In reality, very few legitimate users are usually aware of their biometric being hijacked. This is because our voices can be easily recorded anytime we speak to people. Similarly our fingerprints are retained every time we touch things. This non-secrecy of biometrics, particularly voice biometrics in a speaker verification system, is the reason why retaining the integrity needs to be a primary focus. This non-secrecy characteristic of biometrics emphasizes the need for trustworthy biometric authenticating systems that affirm that the presented voice sample is uttered at the time of verification and not a previous recording; a scenario known as the liveness property [18]. As illustrated in this subsection, the security of any system can not and should not be based on the knowledge of biometric characteristics. Rather it should be based on its ability to adequately address the liveness property.

D. Liveness property

The liveness property confirms that the biometric captured by the sensor device is an actual measurement from the authorized living person. Furthermore, this authorized person should be present at the time of capture. The property aims to guard against the use of dismembered body parts, artificial biometrics or recorded utterances at the sensor device. The testing for liveness is based on either one or more of the following detections: [19]

- Recognition of physiological information as signs of life, e.g. detection of a human pulse using a fingerprint scanner.
- Liveness information inherent to the biometric, e.g. a vein pattern in a hand-geometry verifying system.
- Additional processing of information already captured by a sensor input device, e.g. after capturing a fingerprint scan, to verify that the finger's temperature is not below a certain value.
- Acquisition of life signs using additional hardware, e.g. while doing an iris scan, also perform face recognition.
- Challenge/response protocol, e.g. the system asks the user a question and, only on a correct response, the user is granted system access.

Matsumoto [20] found that many commercial biometric authenticating systems are easily fooled by fabricating the biometric. In his work he used "gummy fingers" to fool fingerprint scanners. Since then, numerous scholars have placed considerable emphasis on the importance of the liveness property in biometric authenticating systems. Often the task of verifying the liveness property of a biometric sample is usually delegated to the particular biometric sensor device. This delegation of duty means that the sensor device needs to be fully trusted by the biometric authenticating system to only present legitimate, live and genuine samples for comparisons. The trustworthiness of the sensor device is also affected by its utilization. Placing a secure biometric sensor device in an unprotected and exposed open area without any supervision renders it insecure. However, placing the same device in a protected premise with human supervision ensures the system sensor device is more secure [18].

In the next section, the proposed improvements on the FPSV system are discussed in detail.

III. EXTENDING THE FPSV SYSTEM

This paper proposes a modification of the existing FPSV design with concepts adopted from the network link layer and audio processing environments. The new design is not only robust but it is also optimized to reduce the threat of replay attacks while addressing the liveness property adequately. The starting point is to present the model's design.

A. The model

The design of the extended FPSV system should identify replay attack attempts, distinguish protocol runs and verify the liveness property. We suggest using a protocol run identifier, which is a unique and dynamic tag. It is dynamic in the sense that it is well suited to deal with the nature of voiceprints in FPSV systems thus acts as a voice fingerprint. To achieve this dynamic nature, the design has a built-in functionality that binds each test utterance to its protocol run with a unique tag. The tag's information identifies and distinguishes voice utterances and protocol runs from each other as well as establishes the liveness property of requests before granting access. Another component present in the design is a module specializing in detecting these unique tags, extracting them via a detection mechanisms and decoding the information. The model is illustrated in the Fig3.

A user prompts the systems to verify his identity by entering a unique number, which maps to the claimed identity in the database. Also, predetermined at this stage, is the tag interval spacing. The tag interval spacing represents a time lapse after which the anti-replay attack tag is embedded into the voiceprint. This is passed to the anti-replay attack component as shown in Fig3. The actual test voice sample is passed for pattern matching to the enrolled voice templates in the database. If the tag interval in the voice utterance and the predetermined tag interval mismatch, the anti-replay attack module returns FALSE.

The tag interval in the voice utterance is determined by processing the voiceprint and recording each fingerprint signal in the voiceprint.

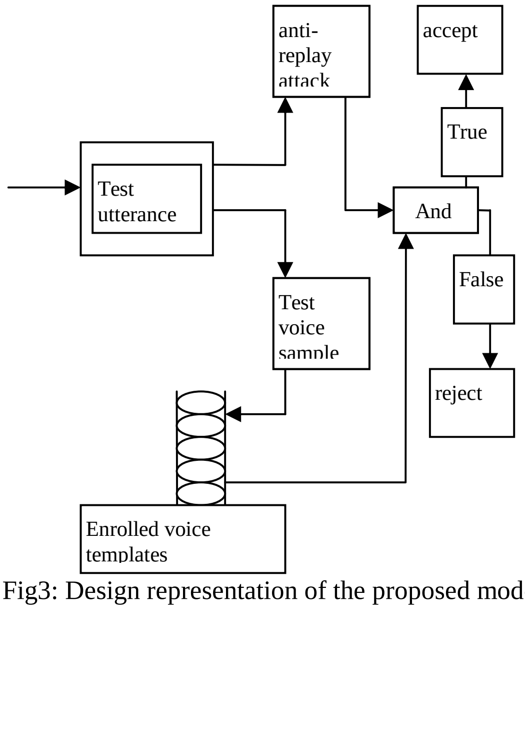

Fig3: Design representation of the proposed model.

This speech fingerprinting technique thrives on the limitations of the human auditory system. However, despite all the imperfection of the human hearing system, there is still enough sensitivity to detect lousy speech fingerprints, this renders the task of designing and implementing good fingerprinting schemes non-trivial. Similarly, if the test voice sample does not match with the enrolled voice templates, a FALSE is returned. These returned values are sent through a logical "AND" operator. Pending the result of this logical operation, a result of a FALSE would mean that the verification process was unsuccessful.

Although the proposed model is still at an infancy stage as such not backed by any quantitative analysis or empirical results, the design concept already suggests a performance gain in that the little training templates are required for the model to function efficiently. This means that a user cannot be prompted to utter a wide range of phrases, unlike the PPSV. The design's separation of concern suggests that the system is upgradeable. The modularized structure further confirms that parallelism has been taken into consideration. In the next subsection, attention is focused on the anti-replay attack module. The anti-replay attack module is responsible for the analysis and decoding of tagged voiceprints to determine if a replay attack attempt exists as well as to verify the liveness property.

B. Anti-replay attack module

The model has a sensor device, fitted with a protected speaker that outputs sound. This sound is produced such that its frequency inaudible to humans without special instruments. This sound as well as other embedded information in the voiceprint represents the anti-replay attack tag or the speech fingerprint. This tag binds to the voice utterance at specific intervals determined on the user first requesting authentication. The binding interval is randomly determined by the systems using Vernam's Cipher process [21]. The binding interval is also simultaneously relayed to the anti-replay attack module. The module then uses this information to verify the liveness property of the speaker as well as determining the positions of the tags for extraction and analysis. The liveness property is verified by comparing the location of the tags in the test utterance with the relayed predetermined interval as well as decoding the embedded information. On discovering inconsistencies; either between the locations of tags in the test utterance and the predetermined interval or speech fingerprint information, then a replay attack is detected. Specifically, the replay attack is one in which a prior protocol run with different tag interval and different speech fingerprint is replayed on the current protocol run. This results in the anti-replay attack module returning false and rejecting the test utterance. Fig3 illustrates this. The block arrow is a secure channel that transmits the predetermined tag interval to the interval comparisons module. The tagged voiceprint is transmitted along the secure channel as indicated by the block arrow. The path indicated by the arrow below the secure channel transmits the voiceprint to the templates database for pattern matching.

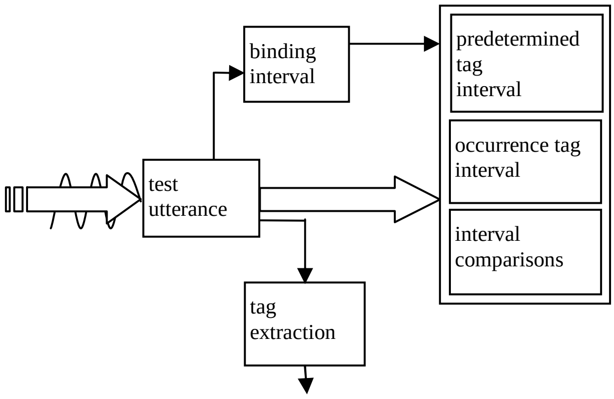

Fig4. Liveness property validation via interval occurrences of tags.

Subsection C delves into the details of determining the tagging interval, tagging the voice print, relaying the tagged voice print, tag interval extraction and the final comparison to find out if the test utterance passes the test.

C. Tag interval determination

The tag interval determination is based on the Vernam cipher [21]. This cipher is a variation of the one time pad; a cipher practically immune to most cryptanalytic attacks. The cipher is immune because it is based on an arbitrary long sequence of non-repeating random numbers and, as such, has no patterns. The lack of patterns in the random number generation means it is practically impossible to predict the next number on which the tag interval relies [21].

In a typical speaker-verification scenario, a claimant ***C*** requests the sensor device to verify his/her identity. He/she enters a code that is stored in the database to map the test voice utterance to the enrolled voice templates. As shown in Fig4, the tag interval is generated and relayed to the interval comparisons module via a secure channel (block arrow). This is the predetermined tag interval $t_{pd.}$ At the time of test utterance, the sensor device binds the test utterance T_u at each interval t_a, with the anti-replay attack tag to get the tagged test utterance $\{T_u, t_a\}$. This tagged test utterance is then passed to the database for matching after which it is passed to the tag interval comparisons module in turn for replay attack analysis. The sample passed to the database is first passed through a tag extraction module that transforms $\{T_u, t_a\}$ back to T_u prior to matching it in the database. The sample passed to the tag interval comparison module is illustrated by Fig5, which represents a correctly tagged voiceprint with equal tag spacing.

An actual representation of how the tagged voiceprint would appear in voice is illustrated in Fig6, to help clarify the tagging idea [22]. The upper diagram (voiceprint) as shown in Fig5 has slight darker regions that represent the tag's location in the voiceprint. The lower diagram highlights the spacing interval at the dark regions as visible in the upper diagram.

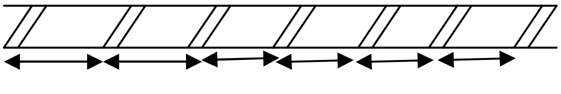

Fig5. Tagged test utterance with the anti-RA tag, of the form $\{\mathbf{T_u, t_a}\}$

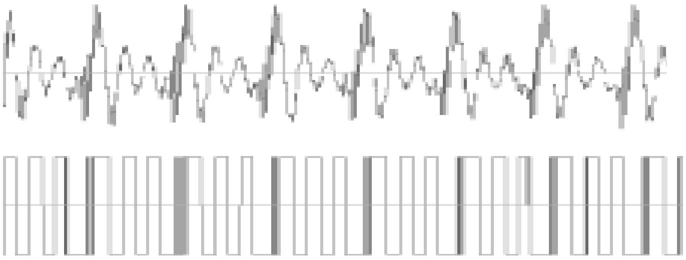

Fig6. Tagged voiceprint tag-interval. [22]

Subsection D details the tag interval extraction from $\{T_u, t_a\}$, as well as the measurement and comparison of the tag intervals from the uttered voiceprint and the predetermined tag interval.

D. Tag interval extraction, measurement and comparison

Tag interval is the distance between anti-RA tags. This distance needs to be extracted from the voiceprint received from the input sensor device and then be compared with the earlier predetermined tag interval t_{pd}. Supposing the supplied test utterance is of the form $\{T_u, t_a\}$, this tagged utterance has a tag interval of t_a microseconds. t_a is computed from the $\{T_u, t_a\}$ using an optimized pitch segmentation algorithm. The extracted value is an indication of the tags' occurrence in the tagged utterance. These occurrences of the tag intervals are then compared with $t_{pd.}$ If the value is such that $t_a \neq t_{pd}$ then it is clear this could be prior protocol run i.e. a replay attack, however, the certainty is confirmed by having $t_a < t_{pd}$. A typical situation described by $t_a < t_{pd}$ is illustrated in Fig7. This module does nothing to verify the user's legitimacy. It only determines if this is a prior protocol run of the system or not.

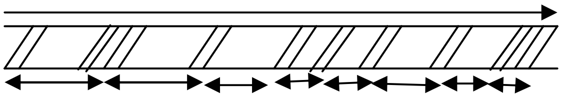

Fig7. Typical example of a replay attack using a prior protocol run, identified by the anti-replay attack module.

IV. DISCUSSION

The design is a fairly straightforward modification of the FPSV model, which is based purely on pattern-matching between an enrolled voice template and a test voice utterance.

The next subsection discusses pivotal issues arising in this paper and current trends. The issues range from biometric thefts, privacy concerns, biometric ethics and the design limitations of the system.

A. Privacy and ethical awareness issues

Currently the major concerns in speaker verification systems relate to the possibilities of a large-scale theft of voice templates and its associated privacy loss. This kind of theft has been in existence for a while, in the form of wiretaps, audio surveillance, telephone taps and recording devices. Today, more than ever, what can be accomplished by stolen voice templates is far greater than what were possible decades ago. One such fact can be attributed to the proliferation of built-in speaker verification systems in most devices e.g. answering machines. Also worth noting is that voiceprints come in varying qualities and are obtainable through using many current tools. Thus, voiceprints used for verification need to be protected. Voiceprints also need to be of high quality. An accurate voiceprint has a similar confidence to a fingerprint or a retina scan. Often underestimated, is the difficulty of a replay attack as the imposter may be induced to speak longer, hence supplying multiple samples of better quality to enable the verification. Thus, possession of a stolen voiceprint does not by itself enable nor guarantee impersonation.

A speaker verification system does limit privacy to some degree, however, this is not out of line with common expectations of privacy. Rather, it borders on the infringement of personal space. A desirable biometric authenticating system would be one where an implicit agreement that one's identity may be ascertained by the participants is negotiated in advance; else one simply opts for a secure session with explicit anonymity.

B. Limitations of the tagged FPSV system

This design model emphasizes on detecting and eliminating replay attacks on which a previous protocol run is replayed. However, it does not discus cases in which voiceprints are recorded out of context and edited before the impersonation. This technique is known as a spliced replay attack [12]. One potentially profound disadvantage is that we have assumed the effect of background and channel noise does not substantially affect the verification performance. In reality this should not be the case. The variability in a legitimate user's voice due to age, disease, emotions and accidents may also affect consistently high recognition results [15]. The current voice distortion, editing and recording devices can also render the speaker verification system impotent, even to legitimate users. Perhaps our greatest limitation is that we may have underestimated the skills of imposters.

V. CONCLUSION

This paper has presented an approach to a FPSV design that reduces replay attacks, particularly those involving previous protocol runs. The binding of an anti-replay attack tag to a test utterance and overlapping the computation of tag analysis with voiceprint matching has resulted in a new FPSV design. The proposed design would not only achieve this at a relatively inexpensive computational level, but it would also address the liveness property at a biometric sensor input point. This has illustrated that the link layer security model can be adapted to implement tagged voiceprints. Thus, with more research on the tags' quality, some more practical use is anticipated and would be supported by means of a prototype. As future research venture tests and experimentations on the of performance will be conducted presented.

The ridge between scholars advocating for improvements in pattern-matching, algorithm optimizations and system design improvements, need to be addressed too. A recent American survey [12] suggests that while about 78% of people are willing to migrate to biometric authenticating systems to perform routine tasks, like automatic teller machine (ATM) transactions, very few of these have any real experience with biometrics, meaning more effort on proliferating the technology is needed [23].

Any future development of speaker verification system technology should take the following into consideration: the difficulty associated with the liveness property should be reduced and one should consider the possibility of encrypting voice prints and still be able to use them for comparisons [23].

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Research Foundation under Grant number 2054024. Any opinion, findings and conclusions or recommendations expressed in this material are those of the author(s) and therefore the NRF does not accept any liability thereto.

[1] REFRENCESMichael Newman, "Speaker verification through large vocabulary continuous speech recognition" Dragon Systems, Inc. 320 Nevada Street, Newton, MA 02160. 1994.

[2] U. Carlsen, "Cryptographic protocol flaws, ". In Proc. IEEE Computer Security Foundations Workshop VII, pages 192–200. IEEE Computer Society Press, June 1994.

[3] James F. Kurose, Keith W. Ross, "Computer Networks, A Top Down Approach Featuring the Internet" Addison Wesley 2nd Edition, 2003.

[4] "Audio Watermarking and Applications" http://xenia.media.mit.edu/~metois/Projects/Waterm/waterm.htm Accessed 23/06/2005.

[5] Hyoung Joong Kim "Audio Watermarking Techniques", Department of Control and Instrumentation Engineering, Kangwon National University, Chunchon 200-701, Korea.

[6] "Cambridge Consultants Ltd", CCL announces a breakthrough in audio watermarking", http://www.cambridgeconsultants.com/news_pr68.shtml. accessed 23/06/2005

[7] Somsak Sukittanon and Les E. Atlas, "Modulation Frequency Features for audio fingerprinting ", Department of Electrical Engineering, University of Washington, Box 352500, Seattle, Washington 98195-2500, USA{ssukitta, atlas}@ee.washington.edu.

[8] Pedro Cano, Eloi Batlle, Emilia G´omez, Leandro de C.T. Gomes and Madeleine Bonnet," Audio Fingerprinting: Concepts and applications". MTG, Universitat Pompeu Fabra, Pg. Circumval.laci´o 8, 08003, Barcelona, Spain. InfoCom-Crip5, Universit´e Ren´e Descartes, 45, rue des Saints-P`eres, 75270, Paris cedex 06, France.

[9] Jaap Haitsma and Ton Kalker, "A Highly Robust Audio Fingerprinting System", Philips Research Prof. Holstlaan 4, 5616 BA, Eindhoven, The Netherlands.

[10] Matthew L. Miller and Manuel Acevedo Rodriguez, " Audio Fingerprinting: Nearest Neighbor Search in High Dimensional Binary Spaces", NEC Research Institute 4 Independence Way Princeton, NJ 08540.

Eurecom Institute BP 193-06904, Sophia-Antipolis, France.

[11] Anil Jain, Lin Hong and Sharath Pankanti, “Biometric identification”, Department of Computer Science and Engineering at Michigan State University., Watson Research Center, Hawthorne, NY

[12] Bruce Schneier, “The Uses and Abuses of Biometrics”, http://www.schneier.com/essay-019.html.

[13] Y. S. Moon, C. C. Leung and K. H. Pun, “Fixed-point GMM-based Speaker Verification over Mobile Embedded System”, Department of Computer Science and Engineering. The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

[14] Robert M. Nickel, “Robust Speaker Verification with Optimal Pitch Bases Expansions,” Electrical Engineering Department. The Pennsylvania State. University.

[15] Lawrence O'Gorman, “Securing Business’s Front Door –Password, Token, and Biometric Authentication”.

[16] Stewart T. Fleming “Biometric Security Concepts, Issues and Flaws”, Department of Computer Science, University of Otago, Dunedin, New Zealand.

[17] Tuomas Aura, “Strategies against Replay Attacks,”, Digital Systems Laboratory, Helsinki University of Technology, Finland.

[18] Zden ek Riha and Vaclav Matyas, “Biometric Authentication Systems”, Faculty of Informatics, Masaryk University.

[19] Stephanie A. C and. Schuckers,”Spoofing and Anti-Spoofing Measures,” Article for Elsevier Information Security Report on Biometrics, Clarkson University and West Virginia University. December 10, 2002.

[20] T Matsumoto, H. Matsumoto, K. Yamada, S. Hoshino, “Impact of Artificial ‘Gummy’ Fingers on Fingerprint Systems”, Proceedings of SPIE, vol. 4677, January, 2002.

[21] Charles P. Pfleeger and Shari Lawrence Pfleeger, “Security in Computing”, Third edition. Prentice Hall, New Jersey, 2003.

[22] Brad Stewart, “low cost voice recognition,” circuit cellar ink, The computer applications Journal.

[23] Lynne Coventry, Antonella De Angeli and Graham Johnson, “Usability and Biometric Verification at the ATM Interface”, Advanced Technology and Research.

[24] Peskin. B et al., “Topic and Speaker Identification via Large Vocabulary Continuous Speech Recognition,” ARPA Workshop on Human Language Technology, Princeton, March 1993.

[25] Nalini K. Ratha, and Andrew Senior, “Automated Biometrics”, IBM Thomas J. Watson Research Center.

[26] Liveness Detection in Biometric Systems, International Biometric Group white paper, Available at http://www.ibgweb.com/reports/public/reports/liveness.html.

[27] Colin Soutar and BioscryptDale Setlak,”Biometric Systems –Threats and Countermeasures –The State-of-the-Art Biometric”.

[28] Gokhan Ttir, Andreas Stolcke, Dilek Hakkani-Ttir and Elizabeth Shriberg, “Integrating Prosodic and Lexical Cues for Automatic Topic Segmentation”. ACM journal.

BIBLIOGRAPHY

Were Oyomno graduated from the University of Nairobi, Kenya with a Bachelors degree in Economics in 2001. He worked for the ministry of lands and Settlement in the Geographical Information System (GIS) department prior to taking up studies at Wits University, South Africa. In 2004, he obtained a Higher Diploma in Computer Science. He is currently studying for his Honours degree in Computer Science at the University Of Pretoria, South Africa.