Multi-Factor Authentication Techniques for Video Applications over the Untrusted Internet

Laith Abbadi

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements

For the degree of

Master of Applied Science degree in Electrical and Computer Engineering

School of Electrical Engineering and Computer Science

University Of Ottawa

Abstract

Designing a completely secure and trusted system is a challenge that still needs to be addressed. Currently, there is no online system that is: (i) easy to use, (ii) easy to deploy, (iii) inexpensive, and (iv) completely secure and trusted. The proposed authentication techniques aim to enhance security and trust for video applications in the untrustworthy online environments. We propose a transparent multimodal biometric authentication (TMBA) for video conferencing applications. The user is identified based on his/her physiological and behavioral biometrics. The technique is based on a 'Steps-Free' method, where the user does not have to perform any specific steps during authentication. The system will authenticate the user in a transparent way. We propose authentication techniques as an additional security layer for various 'user-to-user' and 'user-to-service' systems. For 'user-to-user' video conferencing systems, we propose an authentication and trust establishment procedure to identify users during a video conference. This technique enables users that have never met before to verify the identity of each other, and aims at enhancing the user's trust in each other.

For 'user-to-service' video conferencing systems, we propose a transparent multimodal biometric authentication technique for video banking. The technique can be added to online transaction systems as an additional security layer to enhance the security of online transactions, and to resist against web attacks, malware, and Man-In-The-Browser (MITB) attacks. In order to have a video banking conference between a user and a bank employee, the user has to be logged in to an online banking session. This requires a knowledge-based authentication. Knowledge-based authentication includes a text-based password, the 'Challenge Questions' method, and graphical passwords. We analyzed several graphical password schemes in terms of usability and security factors. A graphical password scheme can be an additional security layer add-on to the proposed multimodal biometric video banking system. The combined techniques provide a multimodal biometric multi-factor continuous authentication system.

Acknowledgement

First, I would like to thank GOD for giving me the chance to pursue my master's degree. I would also like to thank Dr. Carlisle Adams for his feedbacks, suggestions, comments, positive criticism, support, and attitude. This work would not be possible without his encouragement. I would also like to thank my lovely parents, Ihsan Abbadi and Fawzia Abous, for their support, encouragement, and for all the things that could not be described in words. In addition, I would like to thank my dear uncle, Dr. Chris Durso, and my sisters for their encouragement and support. I would also like to thank my colleagues and friends, Ali Noman, Arif Alam, and the rest of Dr. Adams students for their feedback in our monthly meetings.

Table of Contents

Table of Figures

List of Tables

Chapter 1: Introduction

In this chapter, the motivation for this thesis, the problem statement, and a summary of our contributions are discussed.

1.1 Motivation

The internet has turned the world into a small village, and has changed peoples' lifestyles, businesses, organizations, and governments [30]. It enables users to obtain online services, enables businesses to provide online services, and enables governments to provide online services and applications [29]. Nonetheless, it also facilitates attackers and criminals in the organizing and planning of attacks for various objectives, such as for financial, personal, and political reasons [14]. In the past, we have worked at Entrust Inc. and on an IBM funded project. At Entrust Inc, we explored various authentication techniques used by many financial businesses for fraud detection and reduction. At the IBM funded project, we worked on IBM AppScan client side attacks. We analyzed many web attacks and vulnerabilities, and explored new emerging ones.

There is always a need for secure authentication methods for 'user-to-user' and 'user-to-service' systems in order to reduce identity theft and fraud, and to resist against malware and web attacks. These attacks make knowledge-based authentication methods very vulnerable. Furthermore, a challenging problem facing security systems is Man-in-the-Browser attack (MITB). MITB can cause financial fraud [30]. There is no complete solution for MITB; it is a major problem facing online transactions.

There is a need to enhance the security of online authentication methods regardless of location and time, and without adding complexity for the user, and to enhance the user's trust in the interacting parties over the insecure internet. In this

thesis, we propose security and trust techniques for various applications. These include a 'user-to-user' video conference trust establishment procedure for the identification of individuals unknown to each other and the improvement of user's trustworthiness. We also propose a transparent authentication technique for video banking. Furthermore, we analyzed various graphical password schemes that can be used as an add-on security layer for online banking, to address many security concerns faced by the 'Challenge questions' method that is used by various banks.

1.2 Problem Statement

The internet is a place that serves anyone connected to it. Its benefits come with the following drawbacks: incomplete security and trust. Once a device is connected to the internet, it becomes a potential target for viruses, key-logging, screen-capture, spyware, and malware [72]. Furthermore, the internet was designed with no security goals [30]. The following factors aggregate in an insecure and untrusted internet: (i) the availability of computers to anyone, (ii) global connectivity, (iii) the ease by which malware is obtained and distributed, (iv) various motives for the conducting of online attacks, (v) several ongoing and newly emerging attacks and vulnerabilities, (vi) the potential for any system error or vulnerability to lead to potential attacks, (vii) naïve users, and (viii) the availability of targets online [13, 30].

The design of an authentication system that is easy to use, easy to deploy, inexpensive, fully secure and trusted is a challenge that still needs to be addressed. However, various enhancements are necessary to reduce identity theft and fraud, and for resistance against web attacks and malware. To enhance authentication in online environments, we propose authentication techniques for various applications, including 'user-to-user' and 'user-to-service' systems. We propose TMBA, a 'steps-free' system to authenticate an individual during video conferencing. For 'user-to-user' video conferencing, we propose a trust establishment procedure to identify an individual during a video conference. For 'user-to-service' video conferencing systems, we

propose a transparent multimodal biometric authentication technique for video banking. The video banking authentication method can be added to online transaction systems as an additional security layer to enhance the security of online transactions, and to resist web attacks, malware, and Man-In-The-Browser (MITB) attacks. Furthermore, we analyzed various graphical password schemes in terms of usability and security factors. A graphical password scheme can be an add-on to a text-based password serving as a security layer to prevent key-logging attacks.

1.3 Thesis contribution

In this section, we briefly outline our main contributions: (i) a transparent multimodal biometric authentication (TMBA) system; (ii) a 'user-to-user' video conference identification system; (iii) a video banking authentication method; (iv) the analysis of various graphical password schemes and the following conclusion that none of the studied schemes satisfies all usability and security factors; and (v) a multimodal biometric multi-factor authentication system. The main contributions are:

(i) **A transparent multimodal biometric authentication TMBA system based on the fusion of voice and face biometrics.** This system identifies an individual during a video conference, and provides continuous authentication. Furthermore, it authenticates a user in a 'steps free' method, where the user does not have to perform any specific steps during authentication.

(ii) **An end-to-end trust establishment procedure for 'user-to-user' video conferencing.** This technique enables users that have never met before to verify the identity of each other, and enhances the user's trust in each other. It involves the use of a transparent multimodal biometric system that uses the fusion of face and voice biometrics to identify individuals during video conferencing with minimal user interaction. For future 'user-to-user' video conference meetings, a user does not have to perform steps to verify the identity of the other party; the system will verify the identity claim of the other party, and deliver the results to the user.

(iii) **A video banking authentication method.** This proposed method enables users to virtually obtain services, including some which cannot be done through tele-banking or online banking. The method provides a transparent multimodal biometric authentication to identify individuals during a video banking conference. For an online transaction, it is required to guarantee a transaction's security; however, it is not completely guaranteed in the real world. Our method can serve as an add-on to existing techniques, for the securing of online transactions. This add-on security layer can provide continuous authentication and transaction integrity, while decreasing the possibility of harm through malware and MITB attacks. The proposed method provides an approach that is effective for user identification during a video banking conferencing.

(iv) **Various graphical password schemes were analyzed, and it was concluded that none of the studied schemes address all usability and security factors.** The studied graphical password schemes address many security and usability factors. However, none of the studied schemes completely address all usability and security factors. Moreover, through an analysis of various graphical password schemes and the cognitive authentication scheme [130], we have concluded that there is no graphical password scheme that:

(i) is easy to perform without challenging steps,
(ii) is fast to perform,
(iii) is without the use of haptic devices
(iv) requires only simple recall and/or recognizes few pass-images,
(v) has a large password size,
(vi) provides a onetime shoulder surfing resistance, and
(vii) provides full resistance against spyware and many video recorded login sessions.

Moreover, the development of a graphical password scheme capable of satisfying the seven criteria mentioned above is still an open research area.

(v) **A Multimodal biometric multi-factor continuous authentication system.** This system combines knowledge-based authentication with the proposed multimodal biometric system. For a user to employ video banking, the user must be logged in to an online banking session. The online banking system requires the use of a strong knowledge-based authentication method. A graphical password scheme can be used in

addition to a text password. Moreover, the combined techniques provide a multimodal biometric multi-factor continuous authentication system.

1.4 Related Publications

The following are papers based on the work of this thesis.

1. Javadtalab, A.; **Abbadi, L.**; Omidyeganeh, M.; Shirmohammadi, S.; Adams, C.M.; El Saddik, A., "Transparent non-intrusive multimodal biometric system for video conference using the fusion of face and ear recognition," Privacy, Security and Trust (PST), 2011 Ninth Annual International Conference on , pp.87-92, 19-21 July 2011
2. " 'Steps-Free' Multimodal Biometric Method For Video Banking", to be submitted.
3. "A survey of graphical password schemes for online authentication", to be submitted.

1.5 Organization of thesis

This thesis is organized into following sections. Chapter 2 presents a literature review. In Chapter 3, a transparent multimodal biometric authentication (TMBA) system based on the fusion of voice and face biometrics is proposed for user identification during video conferencing. In Chapter 4, a TMBA system for 'user-to-user' video conferencing is proposed. In Chapter 5, a TMBA system for 'user-to-service' video banking is proposed. Also, various graphical password schemes are analyzed in terms of their suitability for employment as an additional security layer to text-based password for online authentication, especially for online banking. Finally, we present conclusions and discuss the directions for future work in Chapter 6. In the

Appendices, various authentication methods are analyzed, and two graphical password schemes are proposed; furthermore ear biometrics fusion for the proposed video banking authentication method is demonstrated.

Chapter 2: Background

There are several applications wherein the proposed techniques can be used as additional security layers for user identification in online environments. These are, for example, online authentication, online banking, 'user-to-user' video conferencing and video banking, among others. The proposed techniques use knowledge- and biometric-based types of authentication. This chapter is thus divided: (i) main types of authentication, (ii) multimodal-biometric authentication, (iii) video conference security, and (iv) multi-factor authentication.

2.1 Authentication Types

The design of an online system must take web attacks and vulnerabilities into consideration, in order to build a secure and trusted system. A key component of a secure online system is a strong authentication method. Appendix A presents authentication methods that are commonly used in online environments. Common authentication types are knowledge-, token- and biometric-based.

2.1.1 Knowledge-Based

2.1.1.1 Text-password and 'Challenge-questions' based

Knowledge-based authentication relies on information known to the user. This can be anything from a personal identification number (PIN) to a text-password, from a question with a challenge-based answer to a graphical password [30]. This technique

requires the user to remember and use the password in question. If the password is forgotten, the user cannot login.

There are many ways to authenticate users in an online environment, and each method has its limitations. In the case of knowledge-based authentication, the limitations would be that the password is liable to be easily forgotten or shared. The *text-based password* is the most used technique of knowledge-based authentication. Studies [30] indicate that humans tend to use text passwords that are easy to remember and that thereby make them vulnerable to guessing attacks. Furthermore, the text-based password is vulnerable to dictionary and key-logging attacks.

The knowledge-based 'challenge questions' method is commonly used by many banks [92], along with text-based authentication. This former method can be frustrating to the user who must remember a set of answers relevant to a list of questions and runs the risk of forgetting them. Furthermore, the answers to 'challenge questions' are often limited to things such as the last name of the user's favourite teacher or the name of the user's pet. Moreover, if an attacker knows a user's information, the answers to 'challenge question' are not exclusive to the user.

2.2.1.2 Graphical Password Schemes

Knowledge-based authentication can also be used with the rapidly-rising use of graphical password schemes. The upcoming *Microsoft Windows 8* operating system will provide its users with the option to log in using a graphical password [106]. Graphical password schemes are intended to address usability and security factors. In terms of usability, the factors to address are (i) ease of use, (ii) time to perform and (iii) ease of remembrance. For security, factors to address are (i) password space, (ii) guessing attack and (iii) observation attacks. The latter observation attacks include (i) shoulder surfing, (ii) eavesdropping, and (iii) spyware including advanced screen capture [123]. A shoulder surfing attack can take place in various ways [37]: (i) an

attacker can observe the user as the password is entered, or (ii) by digitally recording the user when entering the password.

Graphical password schemes are classified into (i) recognition-based, (ii) recall-based and (iii) cued-recall-based [6]. When entering a recognition-based graphical password, the user must recognize previously-chosen images. In the case of recall-based schemes, the user must draw the same password drawn during registration. Finally, when using a cued-recall scheme, the user must click on specific previously-chosen areas of an image [112, 118].

There have been many recognition-based schemes proposed in the last decade. Dhamija et al. proposed the "Déjà Vu" scheme [27], RealUser proposed PassFace [88] and Doja et al. proposed the 'grid-based' [28] graphical password scheme. However, these proposals do not address shoulder surfing attacks. To tackle this problem, Sobrado et al. proposed the 'Convex Hull' scheme [109] wherein the user must click on the area between the pass-objects. This scheme was limited by the delay in login time and by the ease of guessing created by the large convex hull area between the pass-objects.

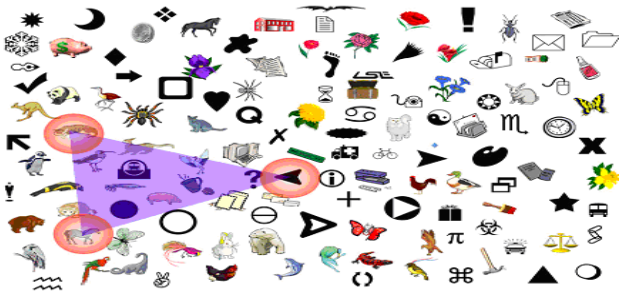


Figure 1: 'Convex Hull' scheme [109]

Wiedenbeck et al. expanded upon the "Convex Hull" technique. In this scheme, the objects are randomly displayed in each round [133]. Moreover, for each round, some decoy-objects are replaced by others and a few pass-objects are added or removed. The area between the pass-objects can be large, and this makes the password easy to guess for an attacker. Furthermore, this scheme is weak against spyware. In every login session, at least three pass-objects are always displayed. If many successful login sessions are observed by an attacker, the attacker can get information about the most appeared objects, which are more likely the pass-objects. Man et al. proposed a

scheme to resist shoulder surfing attacks. Its limitations rested on the user remembering sixteen phrases; a phrase for each of the sixteen icons [71]. Hong et al. extended the Man et al. scheme by having the user create the phrase intended for each selected icon [49]. During authentication, the user would find the pass-objects, beginning from the top-left icon of the screen, and enter the corresponding phrase. The limitation of this scheme is that the user is forced to remember many icons and phrases.

Gao et al. [41] also proposed a 'story-based' scheme to defend against shoulder surfing attacks. During registration, the user selects five images from a set of 'M' images, in order to make a story. The one story is easier to remember, and the user can then also remember the specific, pre-selected order of the images. To recreate the story, the user draws a line that crosses the five pre-selected story images while also crossing images between the pre-selected images [41]. While this technique helps fight shoulder surfing attacks, after some successful login attempts, it is ineffectual against video recording and spyware. Zhao et al. proposed a scheme resistant to shoulder surfing attacks similar to the scheme created by Sobrado et al. This former uses SCII characters instead of objects [139]. Van Oorschot et al. proposed a "two-step" scheme wherein the first step has the user enter a text-based password and then select pre-selected images from a larger set of images in the second step [123]. Weinshall et al. proposed the Cognitive Authentication Scheme [130]. This scheme is limited by the impracticality of the extensive training needed by the user to remember approximately thirty images, especially as there is no safeguard against these images being forgotten over time. The other limitation is that the login requires a few minutes to be completed. This particular scheme was touted as eavesdropping-adverse and spyware-resistant, but this claim was refuted and disproved by Golle et al.: they successfully managed to obtain the password after observing few successful logins [45].

For recall-based graphical passwords, Jermyn et al. proposed the 'Draw-a-Secret' scheme [58] and, basing themselves on the latter, Thorpe et al. proposed a 'selection grid' scheme. Tao furthermore proposed 'Pass-Go': the user would draw a password based on straight lines by connecting grid intersections [115]. In 2011, Microsoft proposed a 'Picture Password' [106]. The upcoming Microsoft Windows 8

operating system will provide users with the option to use the graphical 'Picture Password' to login. At registration, the user will provide a picture and then draw on its surface. During authentication, the system will display the provided picture, prompting the user for the drawing of the password [106]. This drawing is made up of dots and circles, and is intended to connect two or more predefined areas. While this method is easy to use, it still has drawbacks such as shoulder surfing and guessing attacks as well as 'hot-spots'. For touchscreen tablet devices, this scheme is further vulnerable to smudge attacks.

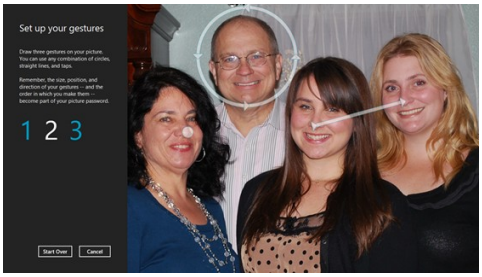


Figure 2: 'Picture Password' scheme [106]

The recall-based schemes, while easy to use, are vulnerable to shoulder surfing attacks. The 'Cued Recall' graphical password schemes, such as, Blonder [7], *Passlogix scheme* [89] and 'PassPoint' schemes [132] are 'repeating a selection' schemes. However, they are vulnerable to 'hotspot' and shoulder surfing attacks.

Graphical password schemes address many security and usability factors. Most of the above-mentioned schemes addressed the issues of brute force searches and guessing attacks, but left the user open to attacks stemming from predictable image selection and 'hot-spots'. Observation attacks including (i) shoulder surfing, (ii) eavesdropping and (iii) spyware still get by the many schemes too weak to counter them. Many schemes were shoulder surfing-resistant only for few login sessions. All of the schemes were ineffectual against spyware and actually indirectly gave the attacker hints about the password.

2.1.2 Token-Based

Token-based authentication is based on something acquired by the user: this can be a paper-based one-time password (OTP), soft tokens, physical tokens or smart-cards. This technique is secure and widely used despite the fact that the user is expected to always have the token on hand to get authenticated [29]. The token can get lost or stolen, its battery can die and maintenance issues can arise with event-based OTP and time-based OTP techniques [29]. The event-based OTP token has issues relating to unused OTP outputs: if and when the user enters a new OTP, the server might be expecting the input of the previous OTP. The issues of the time-based OTP, meanwhile, relate to synchronization with the server: more specifically, the token becomes unusable if it is unsynchronized with the server [29]. The *grid-based* authentication proposed by Entrust Inc. is an add-on to a primary authentication method that depends on an NxM grid-card with numbers and letters (see figure-3). Entrust [31] patented a method of authentication by which the user is asked to enter (i) the text-based password and (ii) the content located at column X and row Y of the user's NxM grid-card. The values of X and Y change with every login. This method is based on a 2-factor authentication. Its drawback, however, is that an attacker can acquire the user's text password and grid content by engaging in shoulder surfing, key-logging or the use of spyware over time.

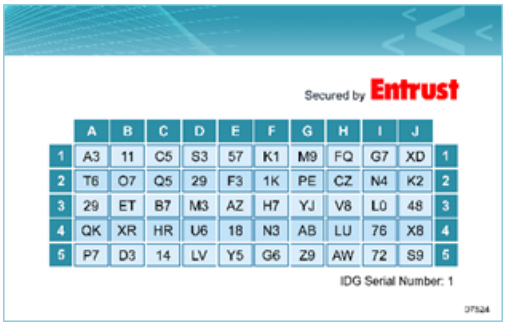


Figure 3: Entrust Inc. 'Grid-Based Authentication' [31]

The *mobile-based* method is often used as an add-on or secondary authentication method to supplement a primary authentication method. For example, Gmail's '2-step verification' requires a user's text-based password to be supplemented by an automatic short message service (SMS) text message or an automated voice message. When the

user enters the text password, the Gmail server sends a message to the user's mobile phone [43]. The user can only get authenticated with a phone that is working and has reception. The *Digital Certificates* method is another method by which to authenticate users. The certificate can be stored on a device, computer, secureID fobs, USB based, or on a smart card-based certificate [30].

2.1.3 Biometric-Based

Jain et al. identified seven features that must be addressed by a biometric system. They include (i) universality, (ii) distinctiveness, (iii) permanence, (iv) collectability, (v) performance, (vi) acceptability and (vii) resistance to circumvention [56]. Universality (i) involves a biometric that must be found in all individuals. Distinctiveness (ii) addresses how unique the biometric is to an individual; compared to the same biometric in other individuals. Permanence (iii) addresses how long the biometric stays unchanged. Collectability (iv) addresses how easy it is to extract features and data from the biometric. Both the accuracy of the output and the recognition rate are addressed in the performance (v). Acceptability (vi) refers to how acceptable the biometric is to individuals. Finally, the seventh feature (vii) relates to how the biometric features are resistant to circumvention. A biometric-based technique can also be divided into two categories, whereas the *biometric-physiological-based* and *biometric-behavioural-based* categories.

2.1.3.1 Biometric-Physiological-based

Biometric-Physiological-based authentication is based on (something the user is) such as: fingerprint, DNA, face, iris, ear, odor, and hand geometry. This method does not require from the user anything other than the user's presence. However, it requires devices and/or scanners, and can be costly. Furthermore, each biometric has its own limitations: in cases of injury, for example, a user's fingerprint can change [55].

Likewise, facial recognition can be deterred by facial hair or by mood and eyes and ears can also be tampered with by the presence of hair or accessories [56]. DNA testing and authentication is just as limited: though it is unique to the user, its use is very expensive [55, 56]. This results in the incomplete accuracy of a biometric-physiological-based system.

2.1.3.2 Biometric-Behavioural-based

Biometric-behavioural-based authentication relies on the user's behaviour. This can include things such as the user's voice, keystrokes, signatures or gait: the only requirement is the user's own behaviour [55]. This method is secure but is limited by the inconsistency of the user's behaviour. For example, a user's voice and signature are easily changed by issues of health, emotion or mood [56]. As such, there is no biometric-behavioural-based system that is completely accurate.

2.2 Multimodal-Biometric Authentication

To avoid the drawbacks of unimodal biometric systems, a multimodal biometric system can instead be used to focus, rather, on circumvention, performance, distinctiveness and accuracy [56]. The fusing of two or more biometrics creates a multimodal biometric system. This fusion, whether pre-classification or post-classification, plays a key role in the performance and in the accuracy of a multimodal biometric system [54].

Pre-classification fusion takes place at the sensor or feature level, before comparison and matching take place [96]. At sensor level fusion, data from different sensors are integrated to create data for feature extraction [56, 75], though data from different sensors cannot always be integrated. Feature level fusion involves different feature vectors from different sensors or from different biometric traits which are then

combined to create a set of feature vectors [54, 55]. However, there are drawbacks. The different sensors or biometrics do not often have integration ability [54]. Commercial biometric traits may not enable or allow access to data at low levels, and the integration of different vectors can lead to a vector that can cause the 'curse of dimensionality' [54, 96].

In post-classification fusion, the comparison and matching have already taken place and the results can thus be used at score, rank or decision level [54, 56]. Score level fusion takes place after each biometric trait outputs its numerical results [17, 54]. The fusion at this level uses a normalization technique on the results. Examples of the former include Min-Max normalization and Z-score normalization. A normalization technique converts results into a common range [0-1], and a common domain (similarity or dissimilarity) domain [48, 54]. In the case of min-max normalization, the estimated minimum and maximum values as well as the set of scores form the normalized scores [17, 54]. The formula is [54]:

$$ns' = (s - min\_s) / (max\_s - min\_s)$$

where ns' is the normalized score, s is matching score, min_s is minimum score and max_s is maximum score. For the Z-score normalization, the normalized score is formed by the arithmetic mean, the standard deviation of the data and the matching score [39, 54]. The formula is [54]:

$$ns' = (s - \mu) / (\sigma)$$

where ns' is the normalized score, μ is the arithmetic mean and σ is the standard deviation of the data. The normalized scores are then combined using any combination rules such as sum rule, simple sum rule, un-weighted sum rule, weighted sum rule, user specific sum rule, product rule, max-rule, min-rule or and product rule [21, 54]. Finally, the total normalized score is then compared to a known threshold value to determine if the identity claim is genuine or if it is an impostor [54].

Rank level fusion takes place when each biometric trait ranks identities in a decreasing order of most to least matching [75, 96]. The fusion at this level can be time-

consuming, especially as each biometric trait matching module has to rank all of the identities in the system [75].

Decision level fusion occurs when each biometric trait outputs its final 'yes/no' decision of whether the user is genuine or an impostor. The fusion can use 'AND' or 'OR' operators, or otherwise make a decision based on the majority [54].

There is much research on multimodal biometric systems [17, 21, 39, 40, 48] using various biometrics. Covavisaruch et al. proposed a multimodal biometric based on the fusion of hand geometry and iris recognition [22]. Fox et al. proposed a system based on the fusion of speech and mouth and face recognition [39] wherein, during authentication, the user must utter a specific statement. As well, Frischholz et al. proposed BioID, a commercial biometric system based on the fusion of face, voice and lip movement [40]. The system requires the user to move the lips. Moreover, Hong et al. proposed a system that fuses face with fingerprint [48]. Chaudhary et al. proposed a multimodal biometric system that is based on the fusion of palm prints, fingerprints and face [17]. There is a lot of research on the fusing of face and ear biometrics and face and voice biometrics [18, 134, 135, 137]. In the case of the former, the user is required to sit in a particular location, pose in a specific position or be seated in the view of a rotating camera so that images of the ear can be captured [135] while in the case of the latter, the user must utter a specific statement [19]. Most of the proposed multimodal biometric systems require the user's presence and involvement.

2.3 Video Conference Authentication and Security

For 'user-to-service' authentication systems, the commonly used authentications are knowledge- and token-based authentication. Rarely do online environments call for biometric-based authentication methods; because such methods require devices to extract and compare the user's biometric features [30]. In the case of a 'user-to-service' video conferencing service, however, adding a biometric-based authentication system as an add-on becomes more feasible. 'User-to-user' identification without third-party

dependence nullifies the possibility of token-based authentication: identifying a user can instead be done using the biometrics- or knowledge-based method of sharing a 'secret'.

Video conferencing is gaining popularity as high-speed internet becomes the norm [16, 102]. This particular method of communication allows users to interact with one another: for example, banks have begun to reach their clients via online video conferences. The security and trust of banking must still exist when it is done in video conference format, and this means that the service's infrastructure must be secure. To enable security for video conferencing, encryption of the audio and video data exchanged between users must be defined [16, 102]. Ensuring a secure video call between end users can be done via the use of (i) a peer-to-peer (user-to-user) network wherein data exchanged get encrypted at the source and decrypted at the destination [102] and via (ii) a client-server based (user-to-server) network wherein the data get encrypted at the source, decrypted at the server, then re-encrypted and sent to the other user [102]. The majority of secure video conference frameworks use standard technologies and protocols [125]. There are many products and research conducted on end-to-end video conference security frameworks [16, 107, 125, 126].

Voice over Internet Protocol (VoIP) uses RTP (Real-time Transfer Protocol) to send data over the internet [86]. If the data are not encrypted, they are vulnerable to eavesdropping and modification [125]. Jitsi, a Java-based, open source application, provides secure VoIP and video calls [125]. To enhance the security of the data exchange over RTP, Jitsi uses Secure Real-Time Transport Protocol (SRTP) and Zfone Real-Time Transport Protocol (ZRTP), an open source protocol developed by Phil Zimmermann, also famed for his creation of Pretty Good Privacy (PGP) [125]. SRTP provides security, integrity and confidentiality while supporting many symmetric, cryptographic algorithms, including AES128 and AES256 [86, 125]. In order to setup an SRTP audio and video session, ZRTP can be used to negotiate and exchange keys, information and parameters [125]. ZRTP does not verify the identity claim: instead, the user must verify the identity of the other user during the video conference. Nefsis [126] is another video conference security framework; it uses encryption and signed certificates [126]. The conference URL starts with "https" and the padlock browser

symbol is present. The SSL encryption resists eavesdropping and packet sniffing [126]. Furthermore, Google Video Chat [16] also secures video conferencing between users. When video calling takes place over 'https', the data exchanged are encrypted between the end user and the server. Google Video Chat is based on internet standards, such as XMPP and H.264 [16]. In addition, VIA3 [102] and Skype [107] use two types of encryption: (i) symmetric encryption through use of AES and (ii) asymmetric encryption (Public Key based) through use of RSA [102, 107]. Skype enables users to have an end-to-end secure delivery of video [107]. Skype-to-Skype video is encrypted using Advanced Encrypted Standard (AES) and is based on digital certificates that provide authenticity, confidentiality and integrity while also protecting against playback attack [107].

There are many available video conference options from which to choose. Parties looking to engage in a video conference need only find and use an application that provides authenticity, confidentiality and integrity. The end user must simply verify the identity of the other user during the video conference, as no application can verify the identity claim.

2.4 Multi-Factor Authentication

Multi-factor authentication provides strong authentication by requiring the user to provide two or more separate proofs of the identity claim [31]. Lavassani et al. presented the concepts of efficiency and effectiveness of common authentication types, as shown in Table 1.

		Costs			Ease of Use	Effectiveness (Security)	Efficiency (O/I)
		Fixed Costs	*Variable Costs*	*Maintenance Costs*			
Knowledge Based		**L**	**L**	**L**	**M**	**L**	**L**
e-Tokens		**H**	**M**	**M**	**L to M**	**M**	**M**
Biometric	**Physiologic**	**H**	**L**	**M**	**H**	**H**	**M-H**
	Behavioral	**L**	**L**	**L**	**H**	**L**	**L**

Table 1: Lavassani et al. table shows the concept of efficiency and effectiveness for authentication types. The weights "L", "M" and "H" represent "Low", "Moderate" and "High", respectively. [62]

The 'effectiveness' presents the security and the accuracy of a system: the higher the accuracy, the more secure the system is [62]. The 'efficiency', meanwhile, reflects a business perspective by presenting the 'cost-benefit analysis' as a ratio of output over input, where the output represents the cost saving by the system and the input is the total cost of the system [62]. Each of the authentication techniques has its benefits as well as its share of weaknesses, limitations and drawbacks. To have a multi-factor authentication, the system has to use at least two types of authentication to authenticate a user. Online authentication methods have to face malware and spyware as well as key-logging and screen capture attacks and these attacks make knowledge-based authentication methods very vulnerable [30]. Another type of authentication is thus required to form a multi-factor authentication [31].

A challenge facing security systems is a Man-in-the-Browser (MITB) attack. As it stands, there is no current solution to MITB: it remains a major problem facing online systems, especially online transactions. Online services, banking and e-transactions require complete security, and must assure they are dealing with the correct user and not an attacker [30, 72]. MITB is an attack by which the browser is controlled by an outside party during online interactions: in this case, the attacker inserts scriptable code in to a Web browser [51]. MITB attacks can occur through (i) browser vulnerabilities, (ii) downloading, (iii) and unintentional downloading of MITB malware [81]. The attacker can modify the user's input without the user's permission or notice [30]. For example, an attacker can modify the 'payment receiver' information and the 'money amount'

value without the user's awareness of the changes. Furthermore, the attack could also change personal information, including the user's mobile phone number [73].

There are various means [30, 72] of preventing MITB. Entrust Inc.'s 'Transaction Verification' sends the user through a 'different channel', the transaction details along with an OTP [29, 30]. The user receives the transaction details and verified the transaction information. From there, the user enters the OTP into the bank's website to confirm the transaction [30]. The 'different channel' is an SMS text message or an automated voice message. Mannan et al. provide a secure transaction system by having a secure mobile application and using a secondary channel. The system assumes that the mobile phone is secure, trusted and malware-free [72]. Entrust Inc. and Mannan et al. provide secure transaction systems [30, 72]. However, using a channel such as a mobile application or mobile-based SMS does not provide complete security. Mobile phones are vulnerable to various attacks and types of malware [79, 59]. If an attacker is able to control the user's computer despite the presence of anti-virus and anti-spyware defenses, the attacker could likely also control the user's mobile [59]. Therefore, these solutions are not completely secure.

Currently, all authentication techniques, including multi-factor authentications, have benefits, weaknesses, limitations and drawbacks. There is no complete security solution that is at once (i) easy to use, (ii) easy to deploy, (iii) low in cost and (iv) preventative of all types of security holes, attacks and vulnerabilities. Moreover, there is no authentication system that is suitable for all types of online applications: each of these must take advantage of what is provided in order to authenticate a user, ideally without having to add devices and requirements. For example, for video conferencing and video banking applications, an authentication system has to take advantage of the user's biometrics presented in the video to identify the user.

Next, we will propose a set of techniques for video conference applications in order to provide an additional security layer, to resist web attacks, malware and MITB attacks and to enhance the security and trust of these online systems.

Chapter 3: Transparent Multimodal Biometric Authentication (TMBA) for Video Conferencing Applications

This chapter is based on a non-intrusive multimodal biometric system that was published in PST2011, and also includes additional work. The paper can be found at [57]. The non-intrusive multimodal biometric system fuses the face and ear biometrics to identify users during a video conference. In this chapter, we propose a transparent multimodal biometric authentication (TMBA) system for video conferencing applications. The proposed TMBA system fuses the user's physiological and behavioral biometrics. The user's face and voice biometrics are fused. The proposed TMBA system uses the voice biometric as a replacement for the ear biometric for its better accuracy rates, higher GAR and lower FAR rates. The proposed TMBA system provides multimodal biometric continuous authentication. Furthermore, the proposed system authenticates a user in a 'steps free' method, where the user does not have to perform any specific steps during authentication. The system aims to be user friendly by minimizing the user's interaction during authentication. The user's presence and basic interaction with the other party is only required during a video conference. This work was developed in a group. Arif Alam helped in conducting an experiment for face recognition. The rest of the work is the author's.

3.1 Introduction

A biometric system is a secure method that can be used for user authentication; however it has several drawbacks. Multimodal biometric systems can enhance the following drawbacks: distinctiveness, circumvention, performance, and accuracy of unimodal biometrics [56]. In the proposed system, authentication occurs by fusing the user's physiological and behavioral biometrics.

The main goals of the proposed system are to:

- Provide more information about the identity of the user during a video conference.
- Automatically identify a user based on his/her physiological and behavioral biometrics.
- Authenticate users in a transparent, 'steps-free' method. The authentication process does not require the user's collaboration (minimal user interaction).
    - The system aims to be user friendly by minimizing the user's interaction during authentication. The user's presence and basic interaction with other party is only required during authentication, unlike other authentication methods such as knowledge-based, token based, and some biometric based such as, fingerprints, an iris scan, a signature, and keystrokes. These all require a user to perform a specific action or step in order to be authenticated.
- Provide a multimodal biometric continuous authentication system

Chapter 3 is organized as follows: Section 3.2 presents the proposed system. Section 3.3 provides the experiment and results. Finally, conclusions are discussed in Section 3.4.

3.2 Proposed System

During a video conference, the proposed TMBA system identifies a user based on his/her biometrical features. The system fuses the user's face and voice biometrics. Based on the fusion of the user's biometrical features, the system will form a decision to indicate if the claimed user is genuine or an imposter. Figure-4 presents a scenario of how the system can be used. The conference is between two users, A and B. Using the system, users A and B will be able to identify one another.

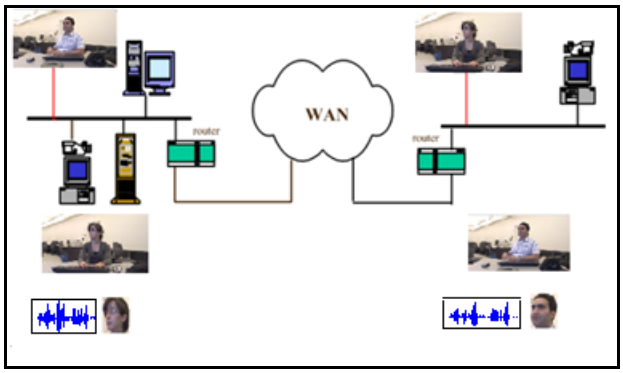


Figure 4: Proposed Transparent Multimodal Biometric Authentication TMBA System [57]

While the users introduce themselves to each other at the beginning of the video conferencing session, user A's system will extract user B's face and voice traits from the video and send them to the recognition modules. Similarly user B's system will extract user A's face and voice traits and send them to recognition modules. The system does not store a user's biometric images or voice samples; only the biometric extracted features are stored into the system to be compared with test images or samples. Once at the recognition modules, the face images and extracted voice of claimed user B are compared to user B's template and the outputs are sent to the fusion modules. The final decision will determine if claimed user B is likely to be genuine or an imposter; user A will then be informed of the decision.

The proposed TMBA system aims to address (i) ease of use, (ii) ease of deployment, (iii) cost of service, and (iv) security. In terms of ease of use, the system is 'steps-free'; the authentication process only requires the presence of the identity claim. The system only requires the user's presence and a simple video conference interaction.

The user does not have to provide specific steps in order to be authenticated. For the ease of deployment, the system uses the basic infrastructure of video conferencing. What differentiates the proposed system from other systems is that it does not require a mobile or any other device to be carried by the user. The system uses a webcam, which is provided during video conferencing. Webcams are built in most laptops, and many users have them. For cost, the system does not require additional equipment, devices and scanners, with the exception of a webcam which is already used during a video conference. For security, the system aims to authenticate a user, based on his/her biometrical features. The system identifies the user based on his physiological and behavioral biometrics. The system enables the other party to verify that it is communicating with the correct user.

3.2.1 System modules

The proposed system extracts the user's biometric traits from the video and audio, and matches them with the user's templates that are securely stored in the system. Furthermore, the system does not store a user's biometric images or samples; only the biometrically extracted features are stored for comparison with test images or samples. The proposed transparent multimodal biometric system main modules are shown in Figure-5. The output of each matching module is sent to the fusion module. The output of the fusion module is then sent to the decision module, where the final decision will be either yes (a genuine user) or no (an impostor).

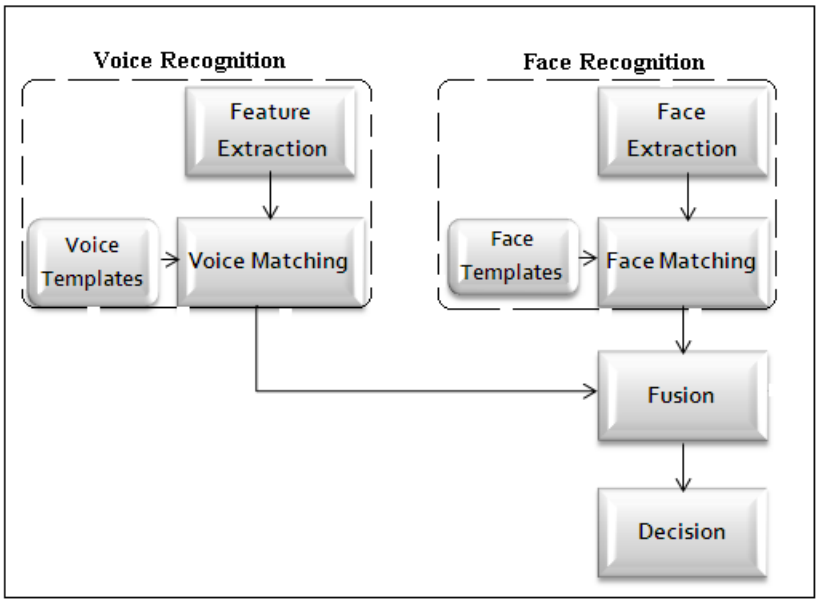


Figure 5: Proposed System.

As illustrated in Figure-5, the modules are Voice Recognition (Feature Extraction Module and Voice Matching Module), Face Recognition (Face Extraction Module and Face Matching Module), Fusion Module and finally Decision Module that outputs the decision regarding whether the identity claim is genuine or an impostor.

3.2.1.1 Voice Recognition (Feature Extraction Module and Voice Matching Module)

Voice, a 'hands-free' biometric is a widely researched topic. Speech recognition recognizes words, while voice recognition recognizes identity [110]. Humans outperform computers in speech recognition. On the other hand, computers outperform humans in voice recognition [95]. There are two types of voice recognition: text-dependent and text-independent. Text-dependent recognition requires the user to utter a word, a random word or number a phrase, or a sentence [110]. However, text-independent recognition does not require the user to utter a specific word, random words, numbers phrase, or a sentence [11, 110]. In this work, we apply text-independent voice recognition.

Text-independent voice recognition has two stages, the training stage and recognition stage [19]. During the training stage, the voice recognition system will take voice samples from the user's voice in order to create a codebook for the user [33]. During recognition, the user's voice is sent to the feature extraction module, and then to the feature matching module in order to output the match score to the fusion module [19]. Feature extraction extracts the voice and converts it to feature vectors [44]. The speech signal is divided into frames. Each frame size is around 20-30 msec [44]. To avoid information loss due to windowing, the adjacent frames overlap each other by about 30-50% [44]. Furthermore, to avoid truncation of the signal, each frame is then multiplied with the generalized window function w(n) [33, 111]. The Generalized Hamming Window formula is [111]:

$$w(n) = (1 - \alpha) - \alpha \cos\left(\frac{2\pi n}{N-1}\right)$$

where N is the length of the signal frame, α equals 0.54 (the standard hamming window), and $0 \leq n \leq N-1$.

A widely used feature extractor for the purpose of producing feature vectors is Mel Frequent Cepstrum Coefficients MFCC [33]. MFCC represents speech based on perception [44]. The MFCC Mel Cepstrum is based on the Mel scale [44]. It has linear mapping for frequency < 1000Hz, and logarithmic mapping for frequency > 1000Hz [19, 44]. The units of the scale are 'mels'. To get the frequency in mels (mel scale), the following formula is applied [44]:

$$Mel\,(f) = 2595\,\log_{10}\left(1 + \frac{f}{700}\right)$$

Furthermore, Delta Cepstrum is used to capture slow changes between the frames [44, 111]. Once the feature vectors are created, a feature matching technique can be applied [21, 44]. A feature matching technique compares the claimed user's features with the ones in the template [44]. Feature matching techniques include: GMM (Gaussian Mixture Modeling), HMM (Hidden Markov Modeling), NN (Neutral Networks), DTW (Dynamic Time Warping), and VQ (Vector Quantization) [11, 33,

77]. There are many possible combinations of features extraction and feature matching. From all the possibilities, MFCC works well with VQ [33].

VQ is a feature matching technique that creates a codebook for each identity [44]. The feature vectors are then clustered using K-means algorithm to make a set of code vectors that represents a codebook [33, 44]. The claimed identity voice data is then compared to the user's codebook; where a distance measure based on the average Euclidean distance is computed [33, 44]. The output is then compared to a threshold value that is determined experimentally to decide whether the user is genuine or an impostor [44]. Moreover, the output is also sent to the fusion module to be normalized and fused with other biometrics.

The threshold value can be determined experimentally based on different voice samples. A simple and easy approach that works was presented in [80]. During authentication, the claimed user's voice data is compared to the user's voice data in the template and to 'N' different users' voices data. Our experimentally created formula based on [80] used to obtain a threshold value is:

$$\frac{X^3}{\left(\frac{Y}{N}\right)^2}$$

Where 'X' is the distance between the claimed user's voice and the user's voice data sample, 'Y' is the total distance between the claimed user's voice to each of the N user's voice data samples, and 'N' is the number of a set of users. Applying the formula for different voices, we noticed that if the claimed user is the correct user (genuine), the output is often within the range of [0.5-3.5]. Some of the outputs are within the range of [3.5-4] and are rarely within the range of [4-6]. If the claimed user is an impostor, the output is often in the range of [5-10]; few are in the range of [4-5], some are rarely in the range of [3.5-4], and some are very rarely in the range of [2-3.5]. The threshold values are in the range of [2.5-4].

3.2.1.2 Face Detection and Recognition (Face Extraction Module and Face Recognition Module)

There has been research on face detection [67, 69, 83, 90]. Robust face detection can detect the face even in the presence of various lighting and background conditions [50]. For the proposed system, the open source computer vision (OpenCV) library has been used for face detection. OpenCV provides several object detectors based on the Viola-Jones method [83]. The common cascade files can detect frontal faces, profile faces, and many facial features, such as: noses, right eyes, left eyes, right ears and left ears [82]. For face detection, the detection is based on the Viola-Jones method [127] and the Lienhart-Maydt method [69], that extends Viola-Jones Haar-Like feature sets for object detection. There are many cascade file options to choose for face detection. A suitable one we found is the 'Tree-based 20x20 gentle Adaboost frontal face detector' [82] created by Rainer Lienhart [69].

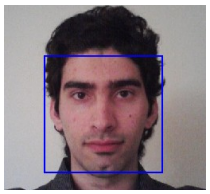


Figure 6: Face detection

Once the frontal face is detected, we apply face recognition. Eigenface is a well-known technique used to recognize human faces. The technique is well defined in [120, 121]. This method projects the training and testing images into a low-dimensional face space [120]. The training process works as follows [121]: from the face images located in the training set M, choose the ones with the highest eigenvalues to form M' eigenfaces, where M' < M [120]. The output defines the face space [121]. The face testing process works as follows; detect the face and extract it. Then project the extracted face into each of the M' eigenfaces, to calculate a set of weights $W_x = \{W_1, W_2, \ldots, W_{M'}\}$ [120]. A classification that is commonly used to measure differences between images is the Euclidian distance technique [121]. Once computation takes place, the distance is compared to an experimentally determined threshold value to conclude whether the claimed identity is genuine or an impostor. Since the face

recognition is part of the multimodal biometric system, the computation results are sent to the fusion module.

3.2.1.4 Fusion Module

Fusion plays a key role in the performance and accuracy of a multimodal biometric system [54]. There are two types of fusion: pre-classification fusion and post classification fusion [19, 54]. For pre-classification fusion, the fusion takes place at a sensor level or at the feature level; before comparison and matching takes place [96]. However, they come with drawbacks [54, 96]. In terms of post-classified fusion, the fusion take place after the comparison and matching process occurs [54]. The three main post-classified fusion are rank level, decision level, and score level [19, 54]. Rank level fusion takes place when each biometric trait ranks the other biometric traits of the same type in a decreasing order from the most matching to the lowest matching [96]. The fusion at this level can be time consuming [54]. Decision level fusion takes place when each biometric trait has already output its final 'Yes/No' decision, whether the user is a genuine or an impostor [19, 54]. The fusion at the decision level can use 'AND' or 'OR' operators; or it may take the majority vote to get the final decision [54].

Score level fusion takes place after each biometric trait outputs its numerical results [17, 54]. The fusion at this level uses a normalization technique on the results. Examples of the former include Min-Max normalization and Z-score normalization. A normalization technique converts results into a common range [0-1], and a common domain (similarity or dissimilarity) domain [17, 54]. The normalized scores are then combined using any of the combination rules, such as: sum rule, simple sum rule, un-weighted sum rule, weighted sum rule, user specific sum rule, product rule, max-rule, min-rule, and product rule [54]. Finally, the total normalized score is then compared to a threshold value to determine if the identity claim is genuine or an impostor [54].

We used score-level fusion described in [54]. By normalizing the results using Min-max normalization technique, and then apply the weighted sum rule.

The min-max normalization is [54]:

$S = Score – min_{Score} / max_{Score} – min_{Score}$

Where Score is the biometric trait score, max_{Score} is maximum score for the biometric trait, min_{Score} is the minimum score for the biometric trait and S is the normalized score.

A normalized score can result in a similarity score or a dissimilarity score [17, 54]. To have a common domain, the biometric trait scores based on dissimilarity scores are converted to similarity scores [17, 54]. To convert dissimilarity score to similarity score, the dissimilarity score is subtracting from 1 [54]. The conversion of each score S results in NS.

The weighted sum rule [17, 21, 97, 129] is then applied:

$Total = W_{B1} *NS_{B1} + W_{B2}*NS_{B2} + … + W_{Bi}*NS_{Bi}$

Where W_{B1} is the weight for the first biometric, W_{B2} is the weight for the second biometric, W_{Bi} is the weight for ith biometric, NS_{B1} is the normalized score for first biometric, NS_{B2} is the normalized score for second biometric and NS_{Bi} is the normalized score for ith biometric.

Each weight W is in range of [0, 1] and the total weight's sum equals to 1 [17, 54].

$W_{B1} + W_{B2} + … + W_{Bi} = 1$

Applying fusion to the system (voice and face) biometrics at the score level, by using the Min-Max Normalization [17, 54] and simple weighted sum rule [17, 21, 97,129], results in;

$S_v = Score_v – min_{vScore} / max_{vScore} – min_{vScore}$

Where $Score_v$ is the voice biometric score, max_{vScore} is the maximum score for voice, min_{vScore} is the minimum score for voice and S_v is the normalized voice score.

$S_f = Score_f – min_{fScore} / max_{fScore} – min_{vScore}$

Where $Score_f$ is the face biometric, max_{fScore} is the maximum score for the face biometric, min_{fScore} is the minimum score for the face biometric, and S_f is the normalized face score.

The normalized score can have a similarity score or a dissimilarity score [17, 54]. The normalized score of the face and voice biometrics are based on the dissimilarity score. To have a common domain, the face biometric score is subtracted from 1 to create NS_f and the voice biometric score is subtracted from 1 to create NS_v.

We then apply the weighted sum rule [17, 21, 97, 129]:

$$NScore_{Total} = W_v * NS_v + W_f * N_f$$

Where $NScore_{Total}$ is the total score from the biometric traits, W_v the is weight for voice biometric, W_f is the weight for face biometric, NS_v is the normalized score for voice biometric, and NS_f is the normalized score for face biometric.

Each of the weights W is in the range of [0, 1] and the total weight's sum equals to 1 [17, 54].

$$W_v + W_f = 1$$

Biometric trait matchers perform differently and can have different recognition and accuracy rates [54]. From the experimental data, we have found that voice biometric outperforms the face biometrics. For this reason, we are putting the weights according to each biometric matcher performance. The voice biometric matcher will have a higher weight than that of the face biometric ($W_v > W_f$). For the fusion, we applied 0.6, 0.4 for W_v and W_f, accordingly. Finally, the fusion module sends the result of $NScore_{Total}$ to the decision module.

3.2.1.5 Decision module

The output of the fusion module '$NScore_{Total}$' is compared to a threshold value to determine if the claimed identity is genuine or an impostor. The threshold value has to

be selected in order to increase the True-Positive (GAR) and to decrease the False-Positive (FAR). The threshold value is within the range [0-1]. To identify the correct user, the threshold value can be from at least 0.5 up to 1. The value depends on the requirements of the system. The higher the threshold value, the less the True-Positive and much less the False-Positive. To have a secure system and to almost eliminate the False-Positives, the threshold value has to be higher than 0.6. For the experiment, we used various threshold values from 0.5-1, as shown in ROC curve in Figure-7.

3.3 Experiment and Results

Jain et al. defined seven features that a biometric system should address. The features include (i) universality, (ii) distinctiveness, (iii) permanence, (iv) collectability, (v) performance, (vi) acceptability and (vii) resistance to circumvention [56]. For (i), each individual should have the biometric. For (ii), the individual's biometric should be unique and different from the biometric in other individuals. For (iii), the biometric should stay unchanged for long time. For (iv), it should be easy to extract features and data from the biometric. For (v), the biometric should have a high recognition and accuracy rates. For (vi), the biometric should be acceptable to individuals. Finally, for (vii), the biometric features should be resistant to circumvention.

For the voice biometric, it satisfies universality, collectability, performance; weak for distinctiveness, permanence, and resistance to circumvention; it is strong for acceptability. For the face biometric, it satisfies permanence, collectability and performance – it is weak for distinctiveness and resistance to circumvention - and is strong for universality and acceptability. When combining voice and face biometrics, the system satisfies collectability, universality, and performance - and is strong for distinctiveness, permanence, acceptability and resistance to circumvention.

3.3.1 Implementation

For face detection, face recognition, voice recognition and cascade detectors, we used OpenCV and the Matlab 7 *open-source code*. For face recognition we used open-source Eigenface implementation. For the voice recognition, we used the Matlab open-source of MFCC-VQ. The database for the experiment was collected between the Summer of 2011 and the Fall of 2011. There are 20 different individual templates containing faces and voice samples. Each template has 3 face samples and 3 voice samples per user. The participating for the face samples are students of University of Ottawa, contractors and employees. Both males and females participating are between the ages of 18 to 45, wherein most of these participants are between 21 to 33 years old. Some users had facial hair and many were wearing eyeglasses. For the voice templates, most of the voice samples are recorded were from various locations that have different noise levels. The sources come mainly from movies, radio and from the University of Ottawa students in an office environment with minimal noise levels. The majority of the users do not speak English as their first language, which can affect the voice recognition accuracy.

3.3.2 Results

A biometric system has four possible outcomes [25, 48]; (i) accepts a genuine (True-Positive); (ii) accepts an impostor (False-Positive); (iii) rejects an impostor (True-Negative) and (iv) rejects a genuine (False-Negative). The correct outcomes are True-Positive and True-Negative and the error outcomes are False-Positive and False-Negative [48]. A biometric system performance can be measured by plotting a Receiver Operating Characteristic (ROC) curve [25, 48]. Wherein the True-Positive represents the standard value Genuine Accept Rate (GAR), the False-Positive represents the standard value False Accept Rate (FAR); the False-Negative represents the standard False Reject Rate (FRR); and, True Negative represents True Reject Rate (TRR) [48]. The standard value (GAR) is the percentage of genuine users getting accepted, (FAR) is

the percentage of impostors getting accepted; (FRR) is the percentage of genuine users getting rejected; and (TRR) is the percentage of impostors getting rejected [25, 48, 75]. The GAR can be determined by subtracting the FRR from 1, (1 – FRR). The curve is plotted in terms of (GAR) versus (FAR) [48] for different threshold (th) values. Figure-7 below plots the ROC curves for the voice and face biometric and for the proposed multimodal biometric system that fuses voice and face biometrics.

For the fused system, we have the following results:
For Th > 0.6, we have 0.9% FAR and 80% GAR acceptance rate.
For Th > 0.55, we have 2.7% FAR and 95% GAR acceptance rate.
The threshold value chosen for the fused system is Th = 0.6.

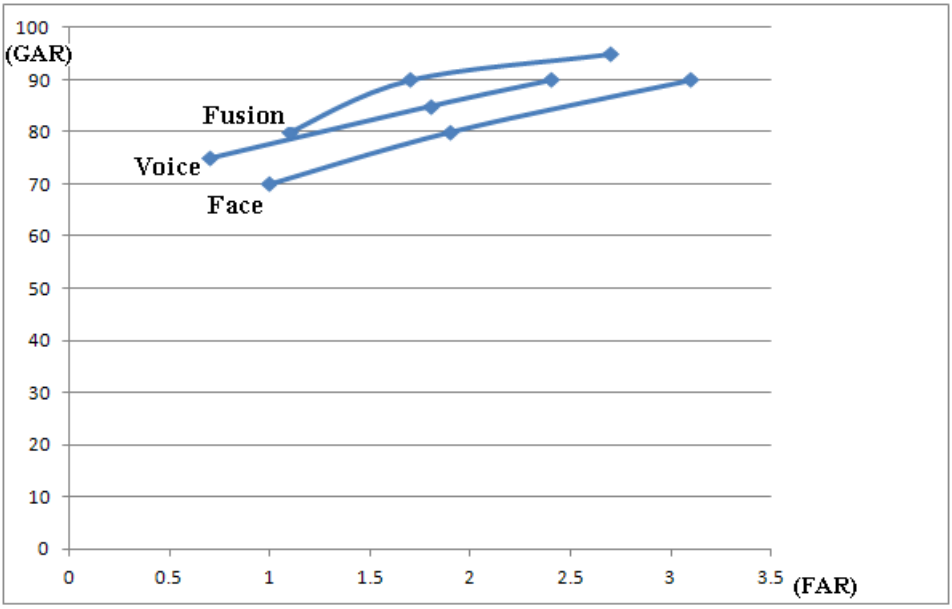


Figure 7: ROC Curves for biometric systems

As shown in Figure-7, the performance of the proposed multimodal biometric system is higher than that of the voice and face unimodal biometrics. However, like any biometric system, the genuine acceptance rate is not 100%. At the decision module, the threshold value has to be chosen so the FAR is around or less than 1% and GAR is at least 80%. For rejecting a genuine (false-negative), the users will not be able to use the system. The aim of the system is to identify users in a transparent method. Alternative methods for authentication can be provided if they cannot become authenticated by

using the proposed system. However, it can require the user to perform steps in a non-transparent way. Also there is a possibility that the system may accept an imposter (false-positive); however, it requires the imposter to look like the user, while possessing a voice close to the user's; this is less likely to occur. Furthermore, the impersonator would get into trouble, since he/she is committing a fraud that is recorded.

3.4 Conclusion

It is a challenge to have an online system that is easy to use, inexpensive, easy to deploy, and completely secure and trusted. However, the proposed TMBA system aims to be easy to use, inexpensive, and easy to deploy, while enhancing security in online environments. The proposed system takes advantage of what is provided without requiring additional devices and equipment. Moreover, the TMBA system provides a transparent multimodal biometric continuous authentication system. In order to have a 'steps-free' user-friendly system, the fusion of voice and face biometrics are used. To authenticate a user, an HD camera and minimal user interaction, during a video conference, is required. The system relies on the existing simple infrastructure of video conferencing and does not require additional devices. The results of the fusion of the multimodal biometrics indicate that the fused multimodal biometric system has a higher performance and accuracy rate than the unimodal biometrics. In the future, the focus will be on using better recognition techniques to enhance the performance of the system by increasing GAR and decreasing FAR. Enhancing the detection and recognition of face and voice biometrics will boost the performance of the system; the aim is to make FAR < 0.1% and for GAR to reach 100%.

Chapter 4: TMBA in 'User-to-User' Applications

This chapter proposes a trust establishment procedure for a transparent multimodal biometric system to identify unknown users to each other during a 'user-to-user' video conferencing.

4.1 Introduction

Weak identity claims lead to unreliable authentication of online users [14]. In this section, an end-to-end trust establishment procedure for a transparent multimodal biometric system is proposed to identify individuals during a video conference. Gaining trust over the internet is a great challenge [13]. In an online environment, the identity of a person or a service with which the user is interacting plays an important role in the user's trust decisions [14]. Thus, transactions are at risk since parties or attackers can provide false identities over the insecure internet [13]. The proposed end-to-end trust system can be used as an additional security layer to authenticate the party at the other end, and uses the simple video conferencing infrastructure without the need for additional devices. Furthermore, the system is transparent after the first video conference: the system will identify a user based on the user's biometrics. The proposed end-to-end trust system provides an effective approach to authenticate users to each other and enhances the real user's trust during a video conference.

Having an online video conference interaction between users can enhance trust. The user's trust decision is affected by many factors, including the machine being used (PC), the other party (a user or a service), the application, and the data exchange [13]. Each of the factors has several trust levels: trust completely, trust for specific things, some trust or no trust at all [14]. The identity of the person or service with which the

user is interacting plays a major role in the trust decision [13]. Since it is easy for a party or an attacker to provide a false identity over the internet, such decisions are at risk [5, 13]. Furthermore, studies [122, 98] indicate that fewer user-to-user interactions result in a higher likelihood that users will lie to each other. Moreover, users are more likely to lie via text messaging than in video and face-to-face interactions [122]. The proposed system provides a video conference interaction between users, thus making users less likely to lie.

For 'user-to-service' authentication, the commonly used authentication techniques are knowledge-based and token-based. The biometric-based authentication methods are not used very often in online 'user-to-service' environments, because most of them require additional devices in order to extract the user's biometric features and compare them with a template [30]. For 'user-to-user' identification independent of a third-party, the usage of token-based authentication is impractical. The choice for identifying an individual can be biometric-based or knowledge-based by sharing a 'secret'. The proposed system is based on face and voice biometrics, and, during authentication, the system requires minimal cooperation from the user. The ear biometric can also be added to this system. To have an end-to-end trust established procedure, biometric authentication can be helpful in establishing the trust in a video conference environment. A biometric authentication system is described in chapter 3.

Obtaining trust over the internet is difficult, especially in an open-system environment [14]. The parties are often unknown to each other, and so trust between them cannot exist. There has been research on Trust Negotiation TN [5, 60, 64, 98] that builds trust on the go between parties that are unknown to each other or that have never met before. To establish trust, the parties exchange the required and minimal policies and digital credentials that satisfy the pre-conditions, while protecting the sensitive data [60]. Trust negotiation enhances authorization issues in open-system environments wherein authorization decisions depend on which resources are requested [64]. The proposed system aims to address authentication and identity claim issues in open-system environments, while the authorization issues are addressed by trust negotiation systems.

We cannot completely solve the problem of trust. The proposed technique, however, aims to enhance trust in an online environment. In this chapter, an end-to-end trust establishment procedure is proposed. This chapter is organized as follows: Section 4.2 presents the proposed end-to-end trust establishment procedure for a multimodal biometric video conferencing system, and the conclusion of the proposed work is in section 4.3.

4.2 End-To-End Trust Establishment Procedure for Multimodal biometric Video conferencing

The proposed end-to-end trust establishment procedure for a transparent multimodal biometric system aims to satisfy a set of criteria, including ease of use, ease of deployment, security, and trust. For ease of use, the user has to perform a set of steps before and during a video conference to identify the other party (user B), but only the first time. After the first video conference, the user does not have to repeat the steps in order to verify the identity claim of the other party; instead, the system will verify the identity claim of the other party automatically. The system also does not require users to own a device such as an electronic token in order to authenticate one another. For ease of deployment, the system uses a simple video conference infrastructure and does not require additional devices. For security and trust, the system provides a set of steps in order to verify the identity of a user. It would provide authenticity, integrity and confidentiality in end-to-end video conferencing. Figure-8 presents a scenario where the proposed system can be used. There are two parties communicating via video conferencing. The system relies on the current simple infrastructure of video conferencing. Moreover, the only device required at both ends is an HD camera, typically found as part of video conference equipment.

E-business often takes place between parties that have never met [53]. The trust between the parties might not exist. For example, a customer (user-A) who wants to buy online goods or services can interact with a seller (user-B), the goods or service provider [5]. To enhance e-business interactions, each party has to trust the other to

perform his/her role [53]. The customer (user-A) has to trust that user-B will send the required goods or services. Meanwhile, the seller (user-B) has to trust that user-A will pay for chosen goods or services. To have ideal security and trust, the entire environment has to be secure. This includes devices, machines, operating systems, applications, other parties (a user or a service) and the data exchange [13]. Moreover, the user is required to have installed up-to-date anti-virus, anti-spyware and firewall [5, 53]. The user makes the final trust decision based on the identity of the other party [13, 14]. But, obtaining complete trust is a challenge over the internet, especially in open-system environments [14]. We cannot totally solve the problem of security and trust, but we are aiming to enhance them to satisfy both parties' minimal requirements.

To ensure the security of the video conference session, the parties have to use an application that provides authenticity, confidentiality and integrity. Users can use Skype [124] or other video conferencing tools to have an end-to-end secure delivery of video. Skype-to-Skype video is encrypted using AES and is based on digital certificates that provide authenticity, confidentiality and integrity while protecting against playback [108]. In the proposed system, we assume end-to-end security between two parties by relying on a video-to-video end-to-end security. Furthermore, reputation systems play an important role in trust decisions and can be used to gain more information about the user [14]. However, current reputation systems have several drawbacks though they gained popularity and were adopted by many online market giants such as e-Bay and Amazon [87].

The proposed system works for 'user-to-user' trust to enhance the user's decisions about what, when, where and who to trust. As shown in figure-8, the system parties are user A, web, user B website, the proposed transparent multimodal biometric authentication (TMBA) system and user B. User A has never met nor knows user B. User A browses the web and finds user B. User B's contact information (name, email, address, fax, photo, etc.) can be found on user B's website. Based on the information, user A uses the proposed system to get steps, information and hints on how and where to search the web to find more information about user B. User A then looks up reputation systems, reviews, business sites and social sites for information about user B.

User A's trust decision is affected by basic scanning for user B's reputation, relationships, deals and past experiences. By a simple search, user A gets a better idea about user-B, by obtaining user B's full name, phone, business, address, photos, reviews, reputation, etc.

User A has to match user B's name with found contact information from different sources and match user B's name with the face photos found at different sources. The goal is to connect the name with contact information, and with face photos. From the gathered information about user B's full name, contact information, and face photos, user A should be able to match user-B's name with contact information, user-B's name with face photos and user-B's contact information with face photos of user-B.

After gathering sufficient information about user B, user A can interact with user B. The first interaction between the users can be via phone or email. During the interaction, user A shares a 'secret' with user B, and they set a date to interact via video conferencing. User A then makes a video conference call to user B. Once user B accepts the invitation and the video conference starts, user A provides identification to claimed user B, and asks claimed user B for the 'secret'. Claimed user B sends the 'secret' back to user A. User A verifies whether the 'secret' matches the 'secret' shared with user B. If it matches, claimed user B is user B. Once the identity of user-B is verified, the transparent multimodal biometric authentication TMBA system via video conference extracts user B's biometric traits, including face and voice, and stores them securely in a template for future verification. The system at user A's end stores user B's biometric features in a secure file under user B's template. The following block diagram is the proposed end-to-end trust establishment procedure for a transparent multimodal biometric system.

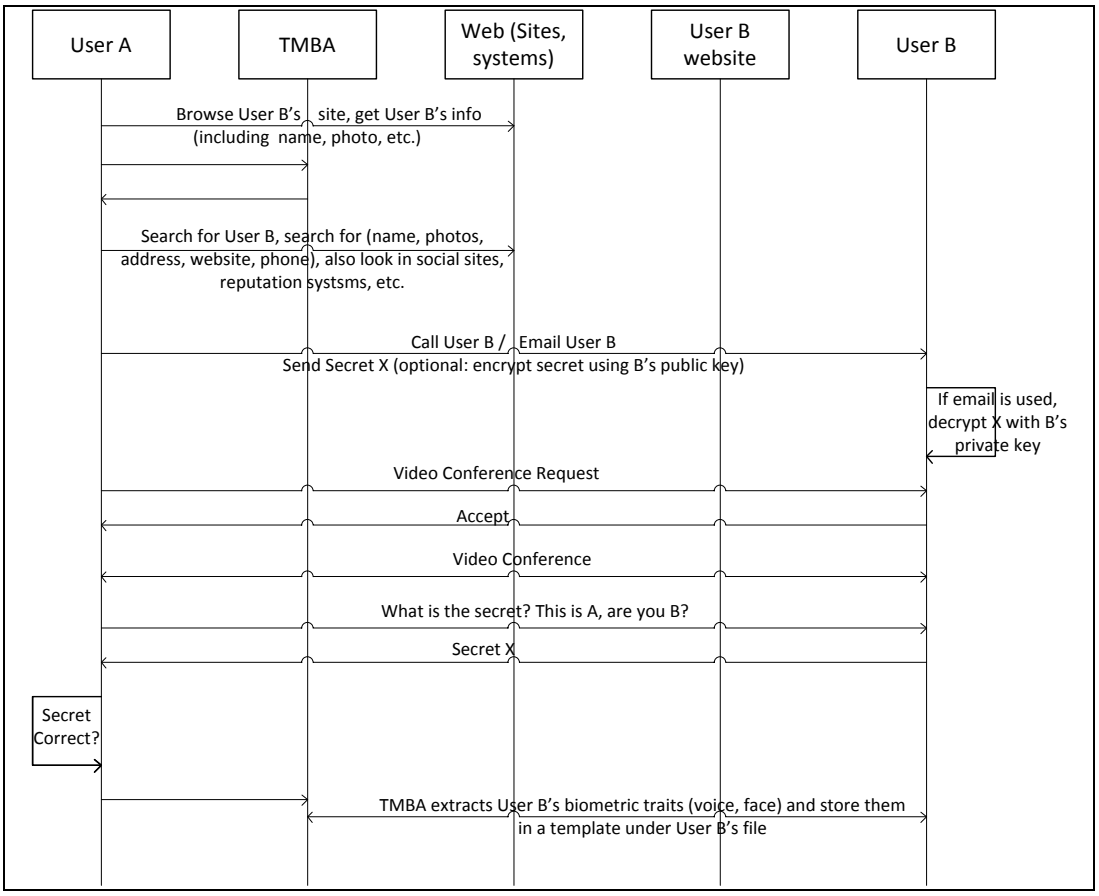


Figure 8: Proposed End to End Security for a Transparent Multimodal Biometric System

In the proposed system, the online environment is not anonymous and each party must reveal its identity. The system can provide traceability, since it is possible to monitor and record video conferences to identify users. However, this raises privacy issues, as users have to reveal themselves to each other.

The following diagram, Figure-9, is the proposed end-to-end trust for a transparent multimodal biometric system after the first video conference.

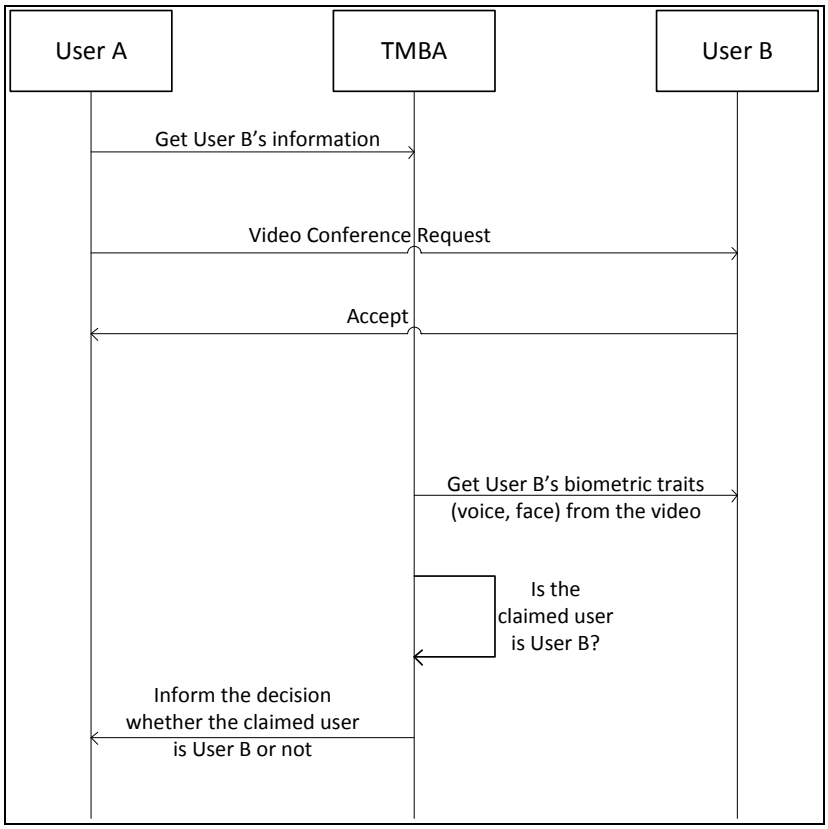


Figure 9: Transparent Multimodal Biometric System after the First Conference

The system at user A's end will, in subsequent video meetings, verify the identity of the claimed user-B at the beginning of the conference, while they both introduce themselves to one another. The system at user A's side will extract user B's biometric traits from the video and match them with the user B templates securely stored at user A's station. The system does not store a user's biometric images or samples; only the biometric extracted features are stored to be compared with test images. Based on the fusion of the biometric traits, the system will form a decision to indicate if the claimed user B is genuine or an imposter. The system will inform user-A of the identity of the claimed user. If the decision indicates that the claimed user-B is genuine, then claimed user B is user B. At this stage, as the trust is enhanced, users can then focus on meeting purposes and negotiate deals and interests.

This system is suitable for users who interact with several parties to do business on a monthly basis. For example, user A, a businessperson, wants to meet user B. During the first meeting, user A had verified the identity claim of user B. For future meetings between the two, user A might find it difficult to verify the identity of user B again without following the same initial set of steps, especially if the former has met with several other people since that first meeting. Therefore, after the first meeting, the user does not have to repeat steps in order to verify the identity claim of the other party; this will be done automatically by the system. Furthermore, the proposed approach enables other users that trust user A, to verify the identity of user B without performing a list of steps. For example, suppose user C wants to verify the identity claim of user B during a video conference. User C knows and trusts user A, but does not know user B. User C requests user B's file from user A. Once user C receives user B's file from user A, he/she can use the system to interact with user B. During the video conference between, user C and user B, user C does not have to perform any steps in order to verify the identity claim of user B: this will be done automatically by the system.

4.3 Conclusion

In this chapter, an end-to-end trust establishment procedure for a transparent multimodal biometric system is proposed. After first time authentication, the system does not require any user cooperation; the user does not have to repeat steps in order to verify the identity claim of the other party as this will be done automatically by the system. The proposed system aims to address identity claim issues in open-system environments rather than the authorization issues of Trust Negotiation (TN) systems. The system is practical, since the solution does not depend on any specific third party. It can be used as an additional security layer to authenticate the identity at the other end. Furthermore, the system uses the existing simple video conferencing infrastructure without additional devices.

To collect information about a user, the proposed system depends on multiple third-parties; however, it does not depend on a specific one. If a third-party is down, the

system still works and relies on the available third-parties. Without fully depending on a specific third-party, it is a challenge to have a complete end-to-end security and trust system for users that are unknown to each other. This area of research is a security and trust challenge that still needs to be addressed. However, the proposed end-to-end identification and trust establishment procedure system provides an approach to authenticate unknown users to each other during a ‘user-to-user’ video conference.

Chapter 5: TMBA in 'User-to-Service' (eg. Video Banking) Applications

Among the services offered by banks, the tele-banking and online-banking services are widely used [94]. In the near future, video-banking will provide more services to users. Video banking aims to enable customers to have additional services not accessible through both tele-banking and online banking [20]. In this chapter, we propose a transparent multimodal biometric authentication method that will serve as an additional security layer for 'user-to-service' video banking. In the proposed method, authentication occurs by fusing the user's physiological and behavioral biometrics. The proposed video banking authentication method provides a multimodal biometric multi-factor continuous authentication. Furthermore, the proposed technique authenticates a user in a 'steps free' method, where the user does not have to perform any specific steps during authentication.

The authentication method aims to be user friendly by minimizing the user's interaction during authentication. The user's presence and basic interaction with a teller is only required during a video banking conference, unlike other authentication methods. The knowledge-based, token based, and some biometric based, such as, fingerprints, scan of the user's iris, a signature, and keystrokes methods, require a user to perform a specific action or step in order to be authenticated [56]. Furthermore, the proposed authentication method can be added as an additional security layer to existing techniques to provide continuous authentication, secure online transactions, and minimize the possibilities of web attacks, malware, and MITB.

Chapter 5 is organized as follows: Section 5.1 is the introduction. Section 5.2 presents the proposed authentication method. Section 5.3 compares the proposed method with other online transaction security techniques. Section 5.4 presents graphical

password schemes add-on to text-based password for online authentication (online banking). The conclusions are discussed in Section 5.5.

5.1 Introduction

Two decades ago, banking services relied heavily on branches and ATM machines [38]. Then the phone based services were added. Later on, internet based online banking was introduced; it has been growing since then [38]. Remote banking services are mainly telephone banking, internet based online banking, and now video banking [94]. Over the telephone, banks often ask users both complicated and impractical questions. For example, to verify a user, the bank may ask a user 'What was your address four years ago?' or 'List your addresses for the last four years?', and etc. An inability to answer the questions fully often results in the user's lack of access to services over the telephone; and the user is asked to go to a bank branch for such services and information.

Video banking is a recent method that provides communication and interaction between parties, such as banks, businesses, and customers via video interaction [94]. In USA, Citizens Financial Group and Charter One Bank launched a video banking pilot in four states [20]. Citizen Financial Group's aim is to provide customers with access to banking when, where, and how they want it [20]. In Canada, FirstOntario Credit Union was the first financial institution to offer video banking [38]. Users interact with bank tellers to do their banking. FirstOntario expanded hours of operation and can be reached by more customers [38]. Video banking aims to be easy, flexible, and cost effective, while enabling customers to have additional services [8, 94]. It provides users with flexibility of location and convenience hours of operation [91, 94]. The service enables users to make remote transactions and to consult an expert or a professional [94].

There are various ways to perform video banking [91, 94]: (i) video banking via enhanced ATM machines; (ii) video banking via video conferencing available through the bank branch, and (iii) video banking via video conferencing through the user's

machine (e.g. PC, laptop, tablet), in addition to the forthcoming HD-TV. These services will be remotely available to customers anywhere and anytime, even after hours of operations [94]. The proposed authentication method works for various video banking and video conferencing sessions. In this work, the focus is on video banking via video conferencing through the user's machine: for example, accessed by computers, laptops, and tablets.

5.2 Proposed Authentication Method

The main goals of the proposed authentication method are to:

- Enable users to obtain online services that are not easily accessible through telephone or online banking.
- Provide banks with more information about the identity of the user in order to provide the user with such services.
    - Provide the service provider with more authentication techniques to authenticate online users.
- Enable the service provider (bank) to automatically identify the user based on his/her physiological and behavioral biometrics.
- Authenticate users in a transparent, 'steps-free' method. The authentication process does not require the user's collaboration (minimal user interaction).
    - The authentication method aims to be user friendly by minimizing the user's interaction during authentication. The user's presence and basic interaction with a teller is only required during authentication.
- Be an add-on for existing online transaction techniques and systems, and function as an additional security layer to minimize online attacks, malware, and MITB attack.

- Be an add-on security layer to assure that it is the user who is performing the online transaction. The parties can use the current transaction security technologies, along with the proposed authentication method to resist web attacks, man-in-the-middle-attack, malware, and MITB attack.

- Provide a multi-factor multimodal biometric continuous authentication system

In order to have a secure video conference for video banking, the infrastructure has to be secure. However, video conferencing uses the insecure internet as a communication channel [38]. To secure a video conference between two parties, the data exchanged including audio and video data must be encrypted [94]. SSL encryption must be used. The framework has to provide authenticity, confidentiality, and integrity.

In the proposed authentication method, the user communicates with the bank through the browser. The communication between the user's browser and the bank's end (website) is based on an SSL connection.

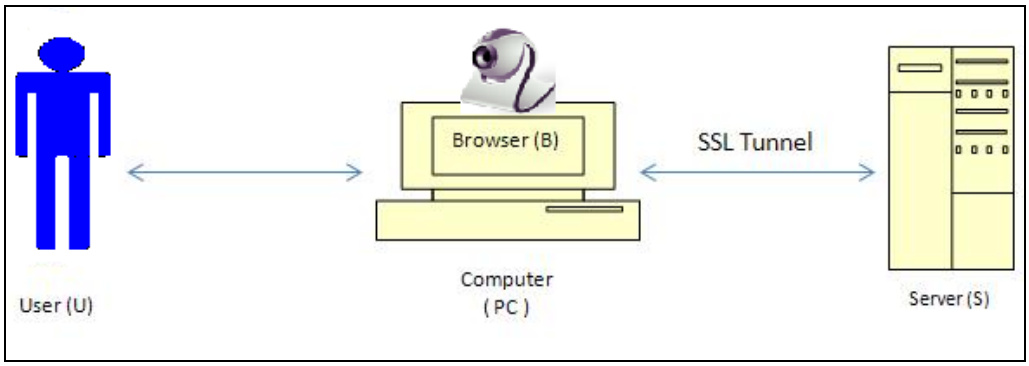


Figure 10: The communication through SSL Tunnel.

The authentication method works as follows:

1. The user U uses a device (e.g. computer) PC
2. U opens a browser B
3. U goes to the bank's website S
4. U logs in to Online Banking
5. U communicates with S through an SSL connection.
6. U launches a Video Banking Application VBA

7. S assigns a bank's teller BT
8. U interact with BT

The main scenario of the authentication method is shown in Figure-11. There are five components of the authentication method: the user, bank, online banking, proposed system, and the bank employee (teller or professional). The proposed authentication method is part of the bank infrastructure and services; it is an integral component to online banking. We assume that the end-to-end video security between two parties is secure, by relying on a secure end-to-end video-to-video application. The proposed authentication method works in the following way: when the user goes to the bank to open an account, the proposed authentication method stores the user's biometric vocal and facial features in a template under the user's electronic file. Once the user uses online banking and requires a 'special request,' the system acknowledges the request, assigns a teller/professional, and initiates a video banking conference between the user and the teller/professional. At the beginning of the video banking conference, the proposed authentication method displays the video to the teller/professional; the proposed authentication method also extracts detected facial images. The voice of the user is extracted and sent to the voice extraction module. The detected facial images are sent to the face matching module. The proposed authentication method then applies a fusion of the biometrics to output a decision, to determine whether the user is genuine or an impostor. Once the identity claim is identified as a genuine user, the authentication method informs the teller/professional to provide services to the genuine user.

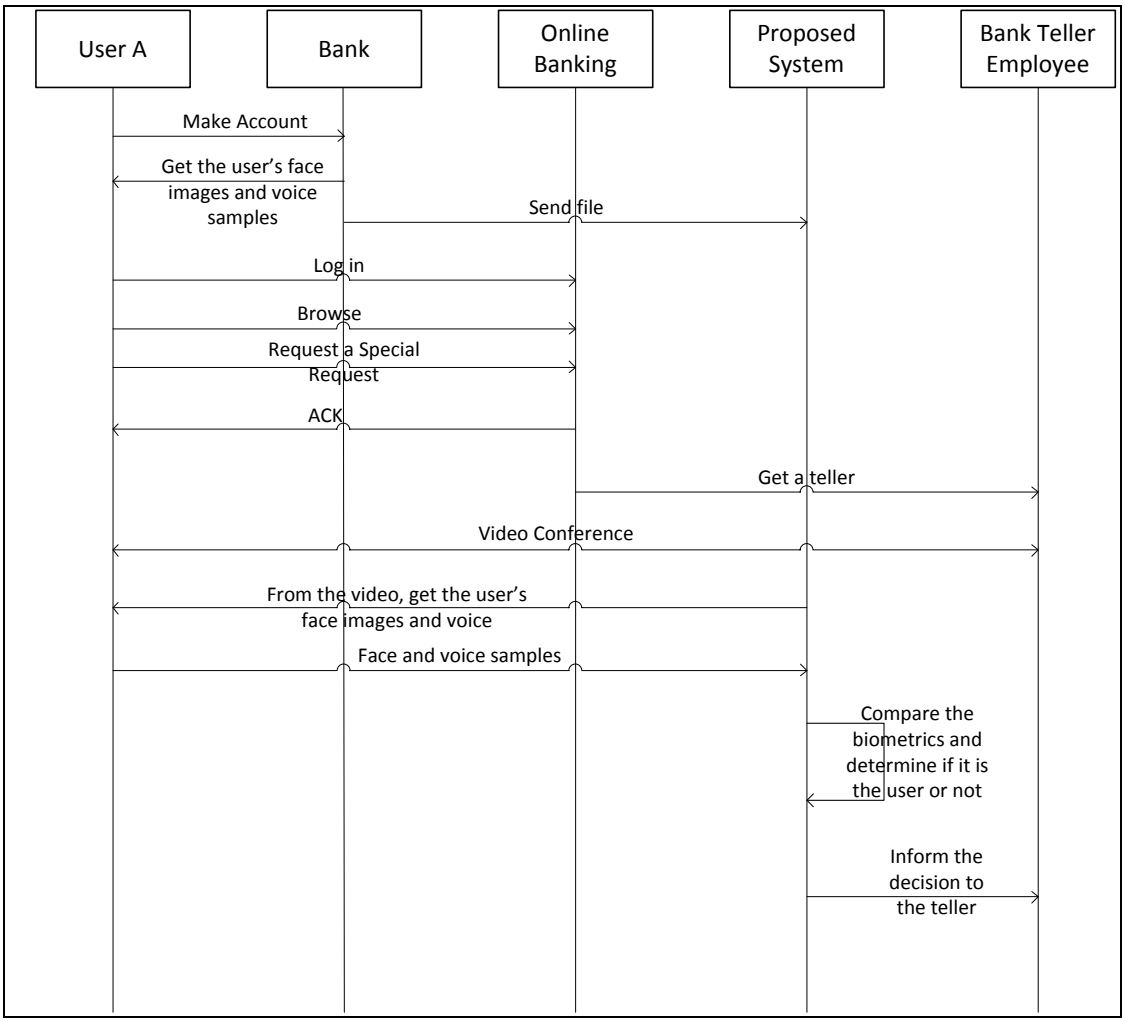


Figure 11: The Proposed Authentication Method For Video Banking

The proposed method aims to address (i) ease of use, (ii) ease of deployment, (iii) cost of service, and (iv) security. In terms of ease of use, the method is 'steps-free'; the authentication process only requires the presence of the identity claim. The method only requires the user's presence and a simple video conference interaction. The user does not have to provide specific steps in order to be authenticated. For the ease of deployment, the method uses the basic infrastructure of video conferencing. What differentiates the proposed method from other systems is that it does not require a mobile or any other device to be carried by the user. And it does not require an installation of software on the user's device/computer and/or on a mobile. The authentication method uses a webcam, which is provided during video conferencing.

Webcams are built in most laptops, and many users have them. For cost, the authentication method does not require additional equipment, devices and scanners, with the exception of a webcam which is already used during video conference.

For security, the authentication method aims to authenticate a user, based on his/her biometrical features, and aims to provide transaction integrity. The authentication method identifies the user based on his physiological and behavioral biometrics. The performance of the proposed TMBA system is higher than that of the voice and face unimodal biometrics. However, like any biometric system, the genuine acceptance rate is not 100%. At the decision module, the threshold value has to be chosen so the FAR is around or less than 1% and GAR is at least 80%. When a genuine user is rejected (false-negative), the user will not be able to use the system. The aim of the system is to identify users in a transparent method and to provide them with services. The service provider can provide users with alternative methods for authentication if they cannot become authenticated by using the proposed system. However, it can require the user to perform steps in a non-transparent way. Also there is a possibility that the system may accept an imposter (false-positive); however, it requires the imposter to be already logged in to online banking. Having the imposter already logged in to online banking represents a serious breach of security. Furthermore, having the imposter already logged in to the correct user's online banking and that the imposter looks like the user, while possessing a voice close to the user's, this is less likely to occur. Furthermore, the impersonator would get into trouble, since he/she is committing a fraud that is recorded.

The authentication method aims to enable the other party, in this case the bank, to verify that it is communicating with the correct user. For online transactions, the authentication method can be added as an additional security layer to verify that the correct user is performing an online transaction. This authentication method is not intended to replace existing online transaction solutions. It can be added as an additional security layer to online transaction security infrastructure current in use, to reduce the possibilities of web attacks, malware, and MITB attacks. Presently, there is no complete solution for MITB. The provided solutions [30, 72] rely on mobile phones and assume

that the ‘second-channel’ is secure, trusted, and malware-free. However, a secondary channel, such as mobile-based SMS, does not necessarily provide complete security.

Online systems often face authentication, confidentiality, integrity, authorization, non-repudiation, and availability threats [128]. To have a completely secure system, an online system has to address authentication, confidentiality, integrity, authorization, non-repudiation, and availability [93]. In our proposed authentication method, we aim to solve security issues that include authentication and integrity; however, we do not claim that the proposed method has solved all the security issues found in online environments. The method does not address confidentiality and privacy threats. It is important to have them; but, for online banking services, we are focusing on authentication and integrity. For confidentiality and privacy threats, if an attacker is able to monitor and control user-A’s PC; then he/she will be able to observe all the information in any case. The method does not address reliability; if an attacker is controlling user-A’s PC, then he/she can stop user-A from having the service. The goal is to provide user authentication and transaction integrity. Furthermore, we assume that the end-to-end video security between two parties is secure, by relying on a secure end-to-end video-to-video application.

During the video banking session, the bank is able to verify that it is communicating with user-A. However, there are possible attacks the attacker can perform to modify the video banking conference; but, these are less likely to occur. In the following subsection, we will identify the possible and potential threats, and indicate how the authentication method can address them.

5.2.1 Threat Model

In this scenario, we have compromised devices (computers). The device that the user is using to communicate with the bank is compromised by malware. The attacker’s goal is to control and modify the communication between the parties to make fraudulent

financial transactions. An attacker can gain control of the browser by controlling the user's device or by MITB attack [72].

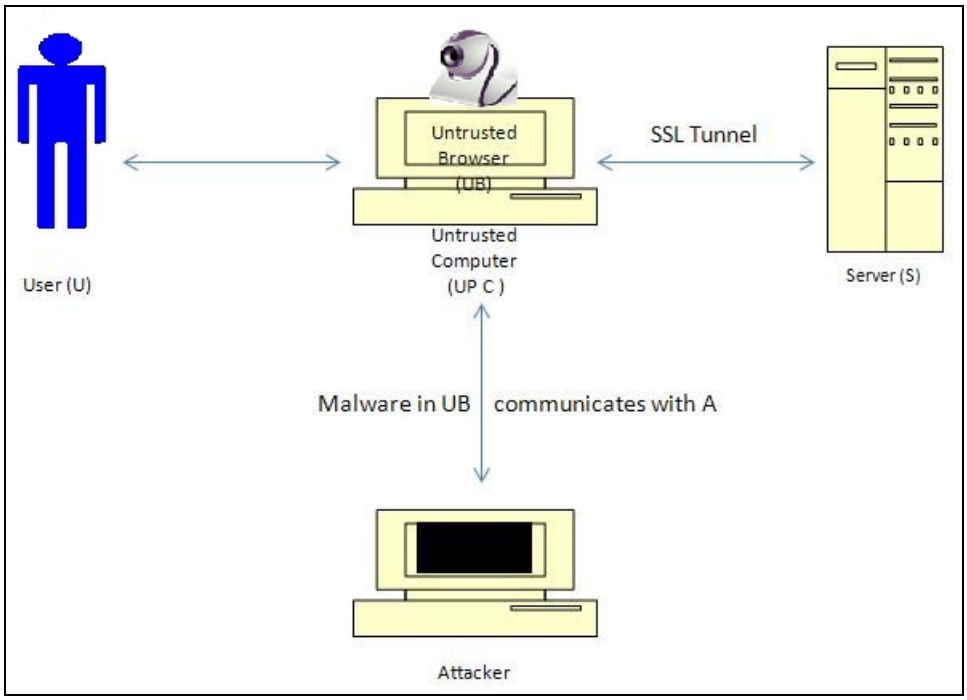


Figure 12: The communication through SSL Tunnel. The user is using a compromised device

The following scenario occurs when the user's device is compromised:

1. The user U uses an untrustworthy device (computer) UPC
2. U opens a browser B
3. U goes to the bank's website S
4. U logs in to Online Banking
5. U communicates with S through an SSL connection.
6. U launches a Video Banking Application VBA
7. Attacker A can control the browser, and can try to:
    a. {view, modify} media content
    b. Redirect communication to the attacker's end instead of bank server end

The following subsections provide unsuccessful attacks against the proposed method and the remaining attacks against proposed method.

5.2.1.1 Unsuccessful attacks against proposed method:

The following are unsuccessful potential attacks against the proposed method.

1) Session hijacking attacks

Session hijacking attacks compromise a session between the user and server in order to gain unauthorized access to the server [84]. The attacker is able to modify or make unauthorized requests and transactions [72]. However, such attacks require the presence of the correct user in order to carry out the unauthorized requests and transactions. If the attacker is controlling user-A's PC, he/she will be able to initiate a video banking session with the bank and impersonate user-A. However, the system at the bank will verify the identity based on the user's biometrics. The system will verify the identity of the user several times during the video session to provide a continuous authentication.

2) Remote desktop attacks

In this attack, the attacker is able to control the user's computer [86]. The attacker can collect the user's input to the bank's website and deny the user access to the website [72]. By controlling the browser, the attacker can login to the bank's website. However, such an attack requires the presence of the correct user to carry out the unauthorized requests and transactions.

3) Attacks against integrity

The attacker can apply media alteration and media injection, where the attacker can inject, delete, and/or replace media content [128]. The attacker has to modify the *contents of the video and audio in real-time*. MITB attack is a method used to modify the integrity [30]. In the MITB attack, the contents of a web form can be controlled and modified [31]. However, in the proposed system, the attacker must have a media tool in order to modify the video and audio content. Furthermore, the attack requires an attacker to have full control over the real time media communication between the user and the bank's end in order to modify media content in real time. The attacker has to

remove some information spoken by user-A, and replace this with the attacker's chosen words. For example, user-A wants to pay '$1,000' (amount) to 'University of Ottawa' (receiver); but, the attacker aims to modify the audio and video of user-A in real time, for the purpose of sending money to a different receiver. When the live conference between user-A and the teller is taking place, the live streaming is performed by user-A's webcam. In order to modify the media, the attacker has to view the communication and act in real time to modify the media content. Therefore, the video stream has to be transferred to the attacker, and not only analyzed but also modified before it is sent it to the teller. Furthermore, the attacker has to produce the same voice as user-A, since the system will verify the voice a second time when the user says the payment and receiver information. In the proposed system the communication is live; and, it is extremely difficult for the attacker to be in full control and to be able to audit all video and audio content between user-A and the teller in real time.

4) Impersonate User-A with an Avatar (type of spoofing attack)

In this attack, the attacker can *impersonate user-A with a 2D/3D avatar*. The attack can occur if the attacker creates a robust 2D/3D avatar that looks exactly like user-A and has the same voice as user-A. Even if the avatar succeeds in passing the verification of face and voice biometrics, it is much more difficult to deceive the bank's teller. Currently, from a video, a human can distinguish a real person from an avatar [76]. In the future, as avatar technologies continue to emerge; further research will be required to digitally distinguish a human from an avatar.

5) Common password attacks

Users often use same passwords for different websites [72]. If the attacker obtains the user's sensitive information including username and password, he/she can start a video banking session. However, such attack requires the presence of the correct user in order to carry out requests and transactions.

6) Phishing and social engineering (threats against social context)

The main threats are phishing, and misrepresentation of identity [128]. If the attacker obtains the user's username and password through key-logging or phishing attacks, he/she can initiate a video banking session with the bank. A phishing attack works through various methods to obtain users sensitive information. However, to carry out an unauthorized request and/or transaction during a video conference, the correct user's presence is required in the communication; where the bank verify the user's identity based on the user's face and voice samples extracted from the media.

5.2.1.2 Remaining attacks against the proposed method

The following are potential attacks the proposed method is limited against.

1) Threats against integrity

a. Media Degrading

The attacker can change the quality of the media by modifying the packets to affect QoS [128]. It only affects the quality of the media. However, an attacker cannot carry out unauthorized requests and/or transactions.

b. Impersonation of the correct user (type of spoofing attack)

One possible attack is to have a user with similar biometrics as user-A to *impersonate user-A*. Since no biometrics can provide complete accuracy, such an attack is possible. To carry out this attack, the user who is the impersonator has to be logged in to user-A's online banking account. Another way to conduct the attack is by having the attacker, in control of the user's browser, replace the video. The attacker can send his/her video to the bank in place of the media streamed from the user's webcam. However, the attacker has resemble user-A or be accompanied by someone who does. Likewise, the attacker or accomplice has to have a very similar voice like user-A.

Furthermore, if the attack occurs, the impersonator will get in trouble for impersonating user-A.

2) Threats against availability

a. Impersonate the server side

This attack is also called 'Media Session Hijacking' [128]. The attacker is able to redirect the communication between the user and the teller to take place between the user and the attacker [128]. The communication is redirected to the attacker's server instead of the bank server. This is a threat involves the misrepresentation of identity, where the user believes he/she is communicating with the bank teller, but is truly communicating with the attacker [93]. The attacker can provide the user with false information and could also gain information from the user [93]. The proposed method does not protect the user from misleading information. However, such threat will not enable the attacker to impersonate user-A in a video conference with the bank's teller. Furthermore, the attacker would not be able to modify contents at the bank's end.

b. Denial of Service Attack / Disabling the communication

The proposed method does not protect against Denial of Service (DOS) attacks. The attacker can provide wrong messages to the server in order to interrupt the communication [128]. Furthermore, when the attacker is in control of user-A's PC, then he/she can stop user-A from obtaining the service. The availability threats can interrupt the service in the form of DOS [128].

3) Threats against confidentiality

Confidentiality threats can arise through eavesdropping media, tracking, and reconstruction [128]. The proposed method does not protect the privacy of the communication. Through the untrusted computer and browser, the attacker could be able to record the communication (media) between the user and the bank's teller. Furthermore, an attacker in control of user-A's PC can capture media and send it to his/her server. An attacker that is able to monitor and control user-A's PC will also be able to observe all the information in any case.

5.3 Comparison with other online transaction security techniques

The main goal of the proposed authentication method is to enable users to obtain online services that are not easily available through telephone or online banking. The proposed method aims to provide banks with more information about the identity of the user in order to provide the user with such services. The proposed method provides the service provider with more authentication techniques for the authentication of online users. This authentication is transparent and provides a 'steps-free' method, where minimal user interaction is required. During authentication, the user does not have to perform a set of steps.

The method can also function as an additional security layer add on to existing online transaction techniques and systems, to minimize online attacks, malware, and MITB attack. The proposed method can provide an add-on security layer to assure that it is the user who is performing the online transaction. The parties can use the current transaction security technologies, along with the proposed method, to resist many web attacks, man-in-the-middle-attack, and MITB attack. Multi-factor authentication does not solve the authenticity and integrity issues of online transactions. For example, if a user logs in with a username and password (knowledge-based) and with an OTP token (token-based); a malware or a MITB attack on a PC/browser can break the authentication system [72]. When the user enters OTP, the malware can take this and quickly log in to the user's account before the user's log in [72]. Our proposed method and the methods in table-2 aim to resist web attacks, malware, and MITB attack.

Table 2: Comparison of the proposed work with other Online Transaction Methods. Note (✓ and ×) represent the optimal scenario. ✓ means the system provides a feature, while × means the system requires a feature.

	Threats addressed			**Enabled**	**Required**	
	Compromised host	Unauthorized transactions (Transaction Integrity)	Keylogging	Privacy	Malware-FREE mobile*	External device
Proposed Method	✓	✓	✓	×		
Mannan Method	✓	✓	✓	×	×	×
Entrust System	✓	✓	✓	×	×	×
IBM ZTIC	✓	✓	✓	×		×

To prevent against MITB, the Entrust Inc. 'Transaction Verification' method [30] sends the user transaction details through a 'different channel', along with an OTP. Once the user receives the transaction details, he/she verifies the transaction information, and then enters the OTP into the bank's website to confirm the transaction [30]. The 'different channel' is an SMS text message. Mannan et al. provides a secure transaction method by having a secure mobile application and also uses a secondary channel [72]. Mannan et al. requires the installation of software on the PC and on the mobile [72]. The method assumes that the mobile phone is secure, trusted, and malware free [72]. Entrust Inc. and Mannan et al. provide secure transaction systems [30, 72]. However, by using a channel such as a mobile application based or mobile based SMS, complete security is not assured. Mobile phones are also vulnerable to various attacks and malwares that computers face [59, 79]. If an attacker is able to control the user's computer that may have anti-virus and anti-spyware software, the attacker could most likely control the user's mobile [59, 79]. Therefore, these solutions are not completely secure.

IBM ZTIC 'Zone Trusted Information Channel' system is based on a second channel concept. The second channel provides 'Out of the browser authentication and verification' [51]. The system is based on an external device required to be carried out by the user. The ZTIC device has a small display screen used to display what the bank provides. The external device has TLS session keys and asymmetric keys stored on it [51]. The data is encrypted with SSL/TLS within the ZTIC [51]. There is direct communication between the external device and the bank's server. Furthermore, the data is encrypted between the device and bank server. The aim is to prevent an attacker from viewing and modifying the data between the external device and the bank's server [51]. However, the main drawbacks of the system are that it increases complexity, and that the use of the system, requires the user to have the physical device.

Our proposed method is based on the infrastructure of video conferencing and only requires a webcam. The method does not require additional devices such as mobile phones or physical devices in order to log in to video banking and to perform online transactions. However, as any biometric system, the proposed method performance is not 100% and requires knowledge-based authentication for online banking before initiating a video conference. What differentiates the proposed method is that it is transparent and is 'steps-free'. During authentication, it does not require a user to perform a set of specific steps to obtain a service or to make online transactions. The proposed method aims to provide a multi-factor multimodal biometric continuous authentication system.

5.4 Graphical password schemes add-on to text-based password for online authentication (online banking)

For added security, the bank may use multifactor authentication (i.e., a login session in addition to TMBA). In order to have a video banking conference between a user and a bank employee, we recommend having the user to be logged in to online banking. The common authentication types to authenticate a user in online banking are knowledge-based or token-based. Token-based authentication can provide a strong

authentication when used with another type of authentication [29]. The token-based drawbacks are that (i) it requires the user to carry the token, (ii) it can be lost, (iii) it can be stolen, (iv) it can malfunction, (v) it can be costly and (vi) it can have maintenance issues for both event- and time-based tokens [30]. Therefore, knowledge-based authentication is widely used to authenticate a user in online authentication. Furthermore, banks [92] often use the 'challenge questions' method in addition to text passwords. The 'challenge questions' method requires the user to answer the questions as were they answered during registration. The drawbacks of this method are that (i) it can be easily forgotten, (ii) it is vulnerable to key-logging attacks and (iii) it is weak against shoulder-surfing attacks. Among the various authentication types, knowledge-based graphical passwords are gaining more attention. The *Microsoft Windows 8* operating system will allow users to log in using a graphical password scheme called 'Picture Password' [106]. Graphical passwords can provide improvement over text-based passwords and 'challenge questions' methods, such as the prevention of key-logging attack. Furthermore, studies [105] show that humans tend to remember pre-selected images and pictures more easily than text.

When designing a graphical password scheme, the design has to address usability and security factors [37]. For the usability factor, a scheme has to address (i) the ease of use, (ii) the time to perform and (iii) the ease of remembrance (how easily a user can recognize or recall the password). For security, factors should be addressed are: (i) password space, (ii) guessing attacks, and (iii) observation attacks [123]. The observation attacks include (i) shoulder surfing, (ii) eavesdropping and (iii) spyware (advanced screen-capture) [112, 123]. A shoulder surfing attack takes place in various ways [37]: (i) by an attacker observing the user while the user is entering his/her password, and (ii) by a user being recorded using an electronic device such as a video camera or a phone. The advanced spyware includes capturing the user's input and the screen content during the entire authentication session [123]. In this chapter, we analyze various graphical password schemes to determine if they are suitable to provide multifactor authentication.

5.4.1 Usability and Security Factors

In this section, we will list the usability and security factors that need to be addressed by graphical password schemes.

5.4.1.1 Usability Factors

A graphical password has to address the following usability factors: (i) ease of use, (ii) time to perform and (iii) ease of remembrance (how easily a user can recognize or recall the password).

Ease of Use

A text-based password is considered easy to do, since it only requires that the user enter the text-based password [123]. On the other hand, graphical password schemes often require the user to perform a couple of instructional steps [6]. This adds a usability complexity to the scheme. Therefore, a graphical password scheme has to be very simple to do, without requiring the user to perform many instructional steps.

Time to Perform

A text-based password is considered fast to perform, as it takes a few seconds to type a password in [123]. On the other hand, graphical password schemes often take more than few seconds [6]. A graphical password scheme cannot exceed a few seconds to be tolerated by the user.

Ease of Remembrance (how easily a user can recognize or recall the password)

A text-based password is not considered easy to remember, since it can often contain random and special characters used to make the password stronger [6]. On the other hand, graphical password schemes often require the user to recognize pre-selected images or recall a previously-created password [27]. Studies [105] show that humans tend to remember pre-selected images and pictures easier than remembering text [6].

However, a graphical password scheme that requires the user to remember more than a few images or reproduce a complicated recall-based graphical password would add memorability complexity to the scheme.

5.4.1.2 Security Factors

A graphical password has to address the following security factors: (i) password space, (ii) guessing attack and (iii) observation attacks [123].

Password Space

The graphical password schemes password space is often compared to text-based password space [6]. A graphical password that has a password space close to text-based graphical password is considered acceptable, and the ones that have way smaller password space than text-based password are vulnerable to potential attacks such as dictionary attacks, brute force search and guessing attacks [6].

Guessing Attacks

For guessing attacks, many graphical password schemes face the same problem of password predictability as text-based passwords [123]. Graphical password schemes can face the problem of 'hot-spots' and dictionary attacks [6]. Research [26, 78] also indicates that users tend to use weak and predictable graphical passwords. More research in the area of graphical passwords predictability is necessary. However, it is more difficult and time-consuming to set up a brute force attack on graphical passwords than it is on text-based passwords [78, 123].

Observation Attack

The observation-based attacks that graphical passwords have to address are: (i) shoulder surfing, (ii) eavesdropping and (iii) spyware [112, 123].

Shoulder Surfing

A shoulder surfing attack takes place in various ways [37]: (i) by an attacker observing the user while the user is entering his/her password and (ii) by a user being recorded using an electronic device such as a video camera or a phone.

Eavesdropping

Eavesdropping enables an attacker to observe all data between the user and the server [37, 123]. To prevent eavesdropping, HTTPS must be used to provide security and confidentiality between the client and the server [123].

Spyware

There are many types and forms of spyware, and new ones are emerging. Without a user's awareness, a spyware can record the user's input from a keyboard, mouse or screen-touch positions for the touch-based devices, and can capture screen contents [37]. The common ways a device gets spyware are through downloading, especially downloading from untrusted and/or unknown websites, and from a virus [49]. Spyware is a serious threat to websites, users and their sensitive information including passwords [37]. Graphical password schemes have to address spyware attacks.

5.4.2 Graphical Password Schemes

Graphical password schemes are classified into (i) recognition-based, (ii) recall-based and (iii) cued-recall-based [37]. For recognition-based schemes, the user has to recognize pre-chosen images. For recall-based schemes, the user has to redraw a password created during registration. For cued-recall, the user has to click on previously selected location(s) on an image [112, 118].

5.4.2.1 Recognition-Based

There have been many recognition-based schemes proposed in the last decade. Dhamija et al. proposed "Déjà vu" scheme. During the login session, the user selects pass-images (pre-selected images) from a set of 'N' random images [27]. The technique is resistant to dictionary attacks but has several drawbacks such as shoulder surfing attacks and spyware.

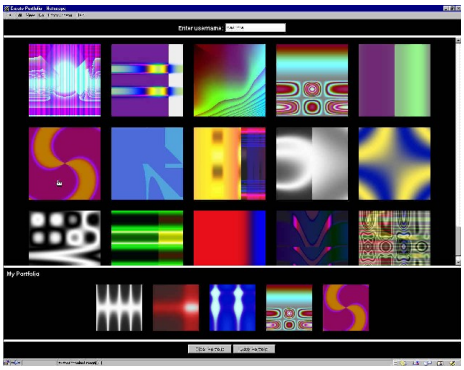


Figure 13: Déjà vu scheme [27]

PassFace [88] is a commercial recognition graphical password. During registration, the user selects 'M' pass-faces from a large database. During the login session, a grid consisting of nine faces is shown, and the user has to click on the pre-selected face. The method is repeated 'r' times. The authors suggested that the value of 'M' pass-faces as 4 and the value of 'r' as 7 rounds. Studies have also shown that users can easily recognize pre-selected faces [88]. The technique is easy to use but comes with drawbacks such as shoulder surfing attack and spyware.

Figure 14: PassFace scheme [88]

Bicakci et al. proposed Graphical Password with Icons, or GPI. In this scheme, the user selects K pass-icons in a pre-ordered sequence from N icons [4]. The scheme drawback is shoulder surfing attack. Alsulaiman et al. proposed a novel 3D scheme [1]. In this, the user has to navigate in a virtual environment to enter the password. The user's textual- or graphical-based password, combined with the user's interaction with 3D objects, create the user's password. Doja et al. proposed a grid-based graphical scheme: during authentication, the scheme displays images on a 12x12 grid square [28]. The user has to choose and click five pre-selected pass-images from the displayed images. The scheme suggests that the pass-images to be selected in a pre-defined sequence. The password space size of this scheme can reach the text-based password space. The technique is, however, vulnerable to spyware and shoulder surfing attacks.

Figure 15: Doja Scheme [28]

To address shoulder surfing attacks, Sobrado et al. proposed a "Convex Hull" graphical password scheme [109]. From the displayed objects, the user has to click on the area between the three pass-objects. From each click, the attacker can obtain hints

about the pass-objects located around the clicked area. Furthermore, the convex hull can be large, making it guessable by an attacker.

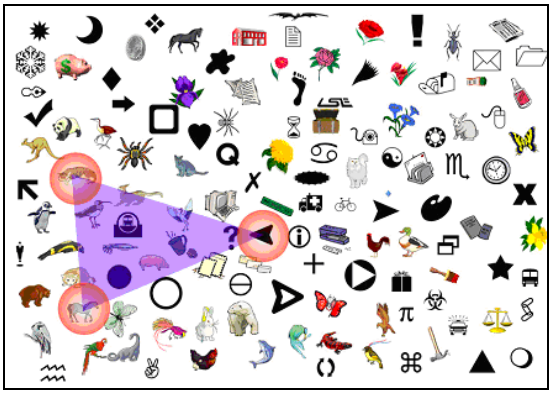


Figure 16: Convex Hull Scheme [109]

Sobrado et al. also proposed the 'pass-objects intersection' technique. In this scheme, the user has to click on the area intersecting four pass-objects [109]. The technique can lead to easy guesses, as the intersection is more likely to be around the centre of the frame. Furthermore, the user must click the expected area in order to login.

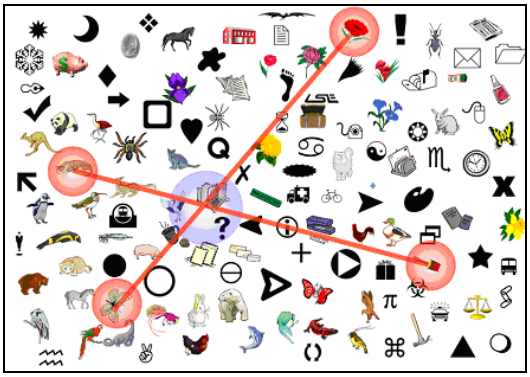


Figure 17: Sobrado 'Object intersections' Scheme. Note that the intersections between the pass-objects are invisible on the screen [109].

Wiedenbeck et al. extended the "Convex Hull" technique. In this scheme, the objects are randomly displayed in each round [133]. For each round, some decoy-objects are replaced by others and a few pass-objects are added or removed. The area between the pass-objects can be large, and this makes it easy to guess by an attacker. Furthermore, this scheme is weak against spyware. In every login session, at least three pass-objects are always displayed. For many successful login sessions observed by an

attacker, the attacker can get information about the most appeared objects and these are more likely the pass-objects.

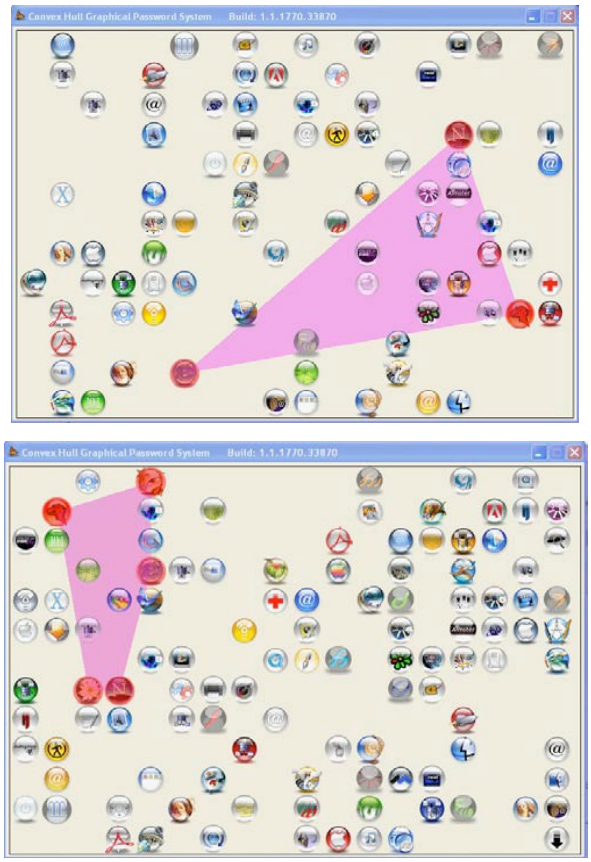


Figure 18: Convex Hull with 3 Pass-Objects, Convex Hull with 5 Pass-Objects. [133]

Man et al. proposed a technique to resist shoulder surfing attacks that requires the user to remember at least 16 phrases, one for each chosen icon [71]. Hong et al. expanded upon the Man et al. technique to resist shoulder surfing [49]. The extended scheme requires the user to remember a phrase that is created by the user for each chosen icon. During authentication, 'N' objects are displayed where there are 'M' pass-objects; the rest are decoy-objects. The user browses the 'N' objects, and, at every pass-object, types the phrase for that object. The major drawback of this technique is that the user has to remember many phrases for the pass-objects.

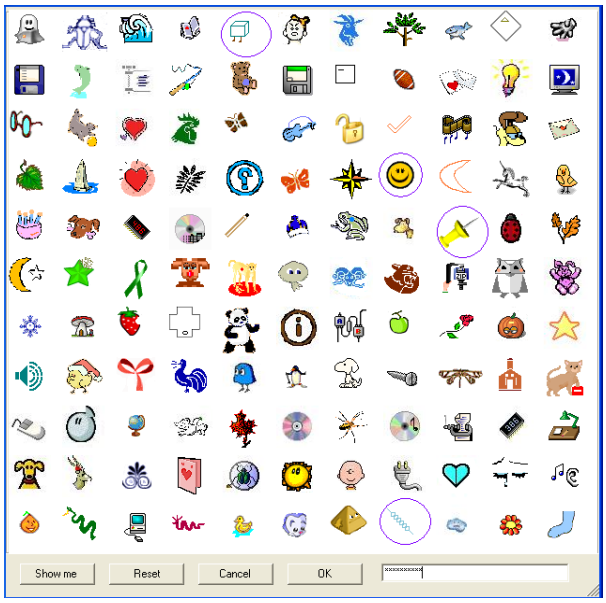


Figure 19: Hong Scheme [49]

Van Oorschot et al. proposed a "two-step" scheme wherein the first step has the user enter a text-based password and then select pre-selected images from a larger set of images in the second step [123]. Each image tops a specific index number. To select a pass-image, instead of clicking, the user takes the index number of the pass-image and clicks on the same number located in the selection panel. The selection panel is displayed under the set of images. In step 2, there are 'r' rounds. This technique has a large password space, since it combines the text-based password with step 2. The drawbacks of this technique are video recording and screen capture spyware.

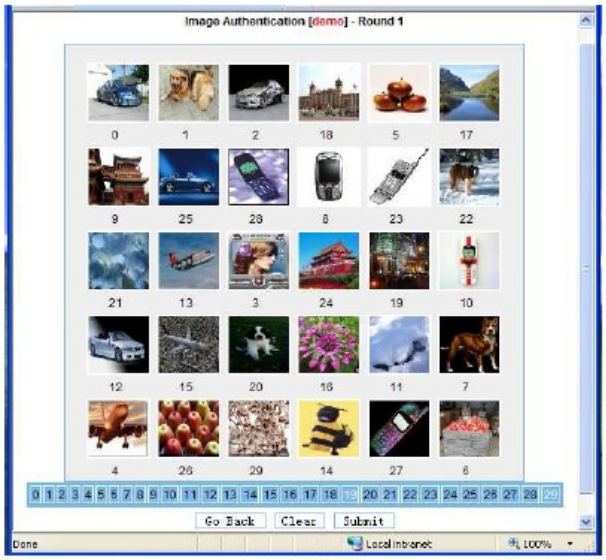


Figure 20: 'Two-Step' scheme [123]

Gao et al. proposed a scheme wherein the user has to recognize pass-images with corresponding colours chosen at registration [41]. During the login session, the user clicks on each row containing a pass-image instead of clicking on the pass-image. The technique aims to resist shoulder surfing attack. However, it is weak against video recorded login sessions. Another drawback is spyware: the observer can get the password within a few recorded successful logins.

(A) First round (B) Completed round

Figure 21: Gao Scheme [41]

Gao et al. also proposed a 'Story-based' scheme to resist shoulder surfing attacks. During registration, the user selects five images from a set of 'M' images, in order to make a story. By remembering the story, the user can recognize the pre-

selected images in specific order. During authentication, the user has to recognize the pre-selected images in order, and then draw a line that crosses them. Other images between the recognized images are also crossed [42]. The technique requires the user to remember five images in a sequence. The technique is weak against video recording and against spyware, as the observer can get the password within very few successful logins.

Figure 22: 'Story-Based' Scheme. In the first frame, the user selects the story images. During authentication, as shown in the second frame, the user creates a path (possible path of password) [42]

Zhao et al. proposed a scheme that integrates text- and graphical-based passwords to provide resistance to shoulder-surfing, video recording and spyware attacks [139]. This scheme is similar to Sobrado et al. [109], but text characters are used instead of objects. For each three pass-characters of the password, the user has to click the character located in the middle of the three pass-characters (see figure-23). For the first click, the user clicks on the character located in the middle of the first three characters of the password. For the second click, the user clicks on the character located in the middle of the second, third and fourth characters. The process is repeated for the rest of the password characters. However, every time the user uses the scheme, a hint is provided to the attacker. Furthermore, using a video-recording or spyware, an attacker observing the login sessions many times will eventually get the password characters that surround the user's click.

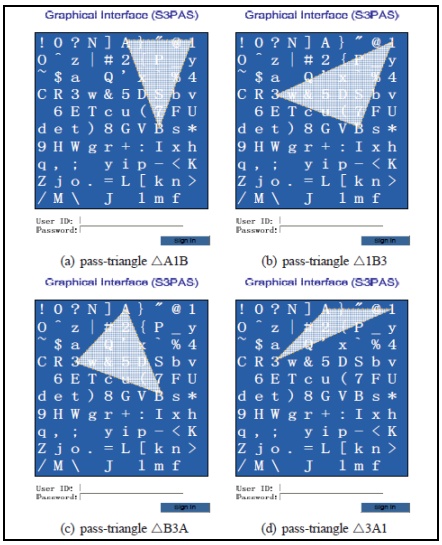


Figure 23: Zhao Scheme. The login for password "A1B3" is the pass characters: P, D, 5, 2. [139]

Weinshall et al. proposed a *Cognitive Authentication Scheme* [130]. In this scheme, there are 'N' images presented in a panel. The user has to compute a path starting from the panel's left-top image. The user moves down if currently standing on a pass-image, otherwise moves right. Once the user computes a path, the number which corresponds to the end point of the panel is entered. There are 'r' rounds. In every round, the user computes a path to get a number. The scheme is resistant to shoulder surfing attacks but requires the user to remember 'k' images out off a set of 'N' images, where 'k' is typically 30 images and 'N' is 80 images. The user has to perform extensive training in order to remember around thirty images. This is impractical, and, over time, the user can forget some of images. Furthermore, the login time is a few minutes. The scheme claims to be eavesdropping-adverse and spyware-resistant. However, the claim was disproved by Golle et al.: they acquired the password in just a few seconds by observing a few successful logins [45].

Figure 24: Weinshall Scheme (Query panel) [130]

In Appendix B, we propose two recognition-based graphical password schemes that aim to resist key-logging and simple shoulder surfing attacks. However, the schemes do not address usability. A security analysis and a formal usability case study are required to determine whether the proposed schemes address security and usability factors.

5.4.2.2 Recall-Based

In recall-based schemes, the user has to recreate something created before [6]. Jermyn et al. proposed the "Draw-a-Secret" (DAS) scheme [58]. In this scheme, the user draws the password on a 2D grid. The technique has a password space larger than an eight-character text-based password [58]. The drawbacks are lack of reliability, possible inaccuracy of the user's drawing, spyware and shoulder surfing attacks.

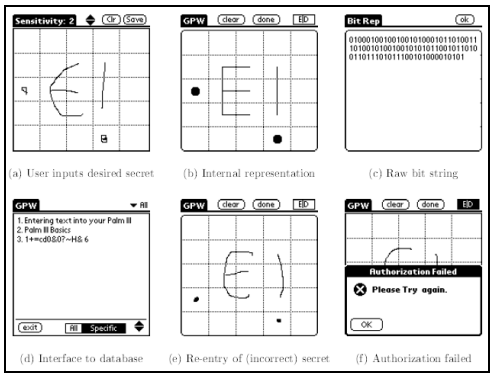


Figure 25: Jermyn's Scheme [58]

Thorpe et al. proposed a selection grid technique. This scheme has larger password size than DAS [119]. The user selects a drawing grid and zooms to draw the secret. The technique shares the same drawbacks of DAS scheme [58].

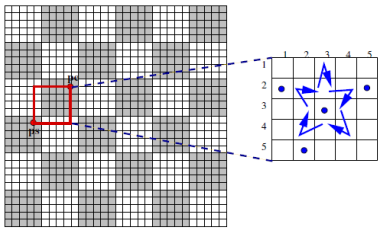


Figure 26: Thorpe Scheme; The user selects a grid to draw password [119]

Tao et al. proposed "Pass-Go". The technique is similar to DAS scheme, but, instead of drawing using grid cells, the user selects grid intersections [115]. The technique provides the user to draw the password more accurately than DAS does. The drawbacks are spyware and shoulder surfing attacks.

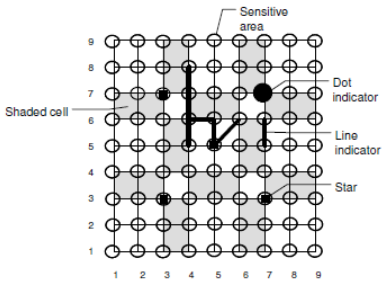


Figure 27: Pass-Go Scheme [115]

Weiss et al. proposed the PassShapes scheme [131]. The user has to makes 7 strokes and each stroke has 7 possible directions. The strokes form the password. The drawbacks are that the password space is small as well as spyware and shoulder surfing attacks.

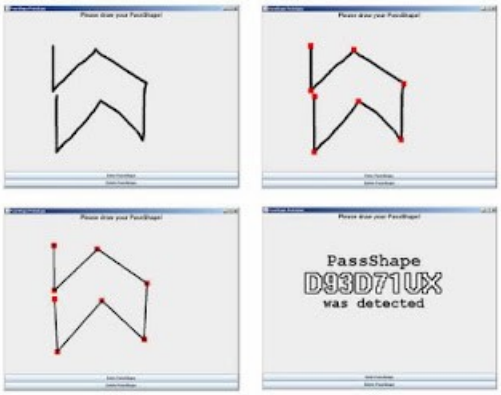


Figure 28: PassShapes scheme [131]

Syukri et al. proposed a scheme whereby the user draws a signature using a mouse [114]. The technique is easy to recall, since it represents the user's signature. The drawbacks include spyware, shoulder surfing attacks and that the user has to use a mouse instead of a pen to draw the password: this can be difficult for some users. Malek et al. proposed a haptic-based graphical password, by drawing a password on a 2D grid [70]. The method is shoulder surfing-resistant since the pen pressure is measured as the user is drawing the password. The user's pressure when drawing the password is used to verify the identity of a user. The experiment shows that such a scheme is guessable, as users apply light pressure when drawing the password.

In 2011, Microsoft proposed a 'Picture Password'. The new *Microsoft Windows 8* operating system will provide users with the option to use the graphical password, 'Picture Password', to login. For registration, the user provides a picture and draws on top of it. During authentication, the system displays the picture on which the user has drawn [106]. The password drawing includes dots and circles as well as connecting two or more predefined areas. The method is easy to use but has drawbacks such as shoulder surfing attack, guessing attack and 'hot-spots'. For tablet devices, the scheme is also vulnerable to smudge attack.

5.4.2.3 Cued-Recall Based

In cued-recall based schemes, the user has to reselect something previously selected (repeat a selection) [6]. Blonder was first to come up with a graphical password scheme in 1996 [7]. In the scheme, the user has to click on several pre-determined locations within an image in a predefined order. The *Passlogix scheme* [89] extended the technique of Blonder. The techniques have drawbacks such as spyware and shoulder surfing attacks. As well as, many pictures and images have 'hot-spots' that are more likely to be clicked than others [118]. This will thus open the door for attacks such as guessing and dictionary attacks.

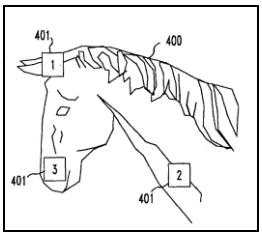


Figure 29: Blonder Scheme [7]

Wiedenbeck et al. proposed 'PassPoint': this expands upon the work of Blonder. It allows a user to click anywhere within an image instead of clicking on predefined locations [132]. The click has to be as close as possible to a pre-selected pixel. To have a larger password space, the user has to perform many clicks in a predefined order. However, the technique is weak against dictionary attacks, spyware and shoulder surfing attacks.

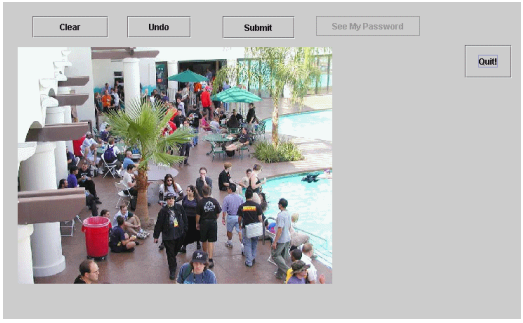


Figure 30: 'PassPoint' scheme [132]

5.4.3 Graphical Password schemes Analysis

Few proposed graphical password schemes address shoulder surfing attacks. Moreover, many that resist shoulder surfing attacks can only do so for few successful login attempts. The majority of the proposed schemes are weak against video recording and spyware (screen-capture) [37]. Many of the schemes give indirect hints about the password. An attacker that records the user's input and screen contents during successful logins will have information about the password and will eventually acquire it.

The following two tables compare various graphical password schemes. The comparisons include the following usability and security factors: (i) ease to do without challenging steps, (ii) time to perform, (iii) the use of external devices and haptic devices, (iv) requires simple recall or recognition of a few pass-images, (v) password size, (vi) resistance to one-time shoulder surfing and (vii) resistance against spyware and video recordings of multiple successful login sessions. The tables can thus determine which of the graphical password schemes satisfy the seven criteria mentioned above. Various graphical password schemes prior to 2006 were analyzed in [112]. The following table analyze various graphical password schemes until 2011. 9 of the following schemes created prior to 2006 were conducted from each graphical password scheme and from [112]. The rest (19 schemes) were conducted from each graphical password scheme.

Table 3: Comparison of text-based, 'Challenge Questions', and graphical password schemes

Techniques	Features	Time to Perform	Password Space*	Drawbacks
Text-based	Fast to perform*, Large password space	Few seconds	92^N (N is the number of password characters)	Hard to remember, dictionary attack, key-logging attack, shoulder surfing attack
'Challenge Questions'	Fast to perform*	Few Seconds	64^N (N is the number of password characters)	Hard to remember, dictionary attack, key logging attack, shoulder surfing attack
Convex Hull Scheme [109]	Shoulder Surfing Resistance (to some extent)*	Similar to Wiedenbeck et al. [13] Approximately 70 seconds. (Mean time to enter correct password) For 5 rounds	$\binom{N}{K}$ Choosing K pass-objects from N objects	Long time to login*, Guessing attack, Hard to remember, Weak against video recording and spyware
Draw a Pass (Haptic) [70]	Fast to perform*, Shoulder Surfing Resistance (to some extent)*	- The user only has to draw the password	$(5^3 \times 5^3 \times 2^3)^n = 50^{3n}$ In a 5x5 grid	Guessing attack
Man [71] & Hong [49]	Shoulder Surfing Resistance (to some extent), Large password size	- Approximately 30 seconds	Approximately same as text based password	Very hard to remember, Weak against video recording and spyware
Association Based [66]	Shoulder Surfing Resistance (to some extend)	For 3 rounds, the average time is less than 30 seconds	$(L \times O \times C / K)^N$ Where: L is the number of the salient points, O is the number of the objects, C is the number of colours, K is the size of the grouped subset of colours, and N is the number of rounds	Guessing attack, Weak against video recording and spyware
Wiedenbeck et al. [133]	Shoulder Surfing Resistance (to some extent)*	71.66 seconds (Mean time to enter correct password) For 5 rounds	$\binom{N}{K}^{\wedge r}$ Choosing K objects from N objects, for r rounds	Long time to login*, Guessing attack, Hard to remember many pass-objects, Weak against video recording and spyware
Blonder [7]	More clicks, the larger the	- Very Simple. The	N^K (N is the number of	Hard to remember, Guessing attack,

	password space	user knows where exactly to click on the image	pixels within image, and K is the number of objects)	Brute force search, Hot-spots attack, Shoulder surfing attack
Dhamija [27]	Resistance to Dictionary Attack	32-36 seconds Average time to login including multiple attempts. Without multiple attempts, less than 30 seconds	$\binom{N}{K}$ Choosing K objects from N	Guessing attack, Shoulder surfing attack
Passfaces [88]	More rounds more password space	14-28 seconds Depending on the number of rounds required	N^R (N is the number of displayed objects, and R is the number of rounds)	Guessing attack, Shoulder surfing attack
Passlogix [89]	Clicking pass-items in a pre-selected order, the more the items the larger the password space	Similar to Pass-point Less than 30 seconds	$\binom{n!}{(n-k)!}$ (n is the number of displayed objects, and k is the number of pre-ordered pass-objects)	Guessing attack, Brute force search, Limited password space, Shoulder surfing attack
DAS [58]	Fast to perform*, Large password space	Recall-based Less than 30 seconds	Larger than text-based password	Shoulder surfing attack
Thorpe [119]	Fast to perform*, Large password space	Similar to DAS Less than 30 seconds	Larger than text-based password	Shoulder surfing attack
S3PAS [139]	Shoulder Surfing Resistance (to some extent)*	Same steps as in Wiedenbeck et al. [13] Approximately 70 seconds (Mean time to enter correct password)	$\binom{N}{K}^{\wedge r}$ Choosing K objects from N objects, for r rounds	Long time to login*, Guessing attack, Weak against video recording and spyware
PassGo [115]	Fast to perform*, Large password space (larger than DAS)	Similar to DAS Less than 30 seconds	Larger than text-based password	Hard to remember (to some extent), Shoulder surfing attack
Syukri [114]	Fast to perform*, Large password space	Similar to DAS Less than 30 seconds	Larger than text-based password	Shoulder surfing attack
Pass-point [132]	Fast to perform*	9 – 20 seconds Average time to login	N^K Where, N = (L x M)/(G x G) (L x M is image size, and (G x G) is grid square size (N is the number of grid squares within	Guessing attack, Brute force search, Hotspots attack, Shoulder surfing attack

			image, and K is the number of areas to be clicked)	
3D [1]	Hard to pass password to others	- Navigate in a virtual environment. The combination of the user's interactions, inputs, and actions forms the user's unique password.	Larger than text-based password	Shoulder surfing attack
NPSRSS [42]	Shoulder Surfing Resistance (to some extent)*	19.8 seconds (Mean time to enter correct password)	$\binom{n!}{(n-k)!}$ (n is the number of displayed images, and k is the number of pre-ordered pass-images)	Very weak against video recording and spyware
'Picture Password' [106]	Fast to perform*, Large password space	Similar to DAS Less than 30 seconds	Larger than text-based password	Guessing attack, Weak against shoulder surfing and spyware
Cognitive Auth Scheme [130]	Shoulder Surfing Resistance (to some extent)* (disproved) [32]	90-120 seconds Average time to login	N^K (N is the number of multiple choice answers, and K is the number of rounds)	Guessing attack, Hard to remember password, Time consuming
DAGPS [41]	Shoulder Surfing Resistance (to some extent)*	13.6 seconds (Mean time to enter correct password) On 12x12 grid	$C\,x\binom{N}{K}$ (C is number of colours, N is the number of icons, and k is the number of chosen icons)	Very weak against video recording and spyware
Recall-based Decoy Strokes [138]	Fast to perform*, Shoulder Surfing Resistance (to some extent)*	Similar to DAS Less than 30 seconds	Approximately same as text based password	Hard to remember (to some extent), Dictionary attack, Very weak against video recording and spyware
Recall-based Disappearing Strokes [138]	Fast to perform*. Shoulder Surfing Resistance (to some extent)*	Similar to DAS Less than 30 seconds	Approximately same as text based password	Hard to remember (to some extent), Dictionary attack, Very weak against video recording and spyware
Recall-based line snaking [138]	Fast to perform*, Shoulder Surfing Resistance (to some extent)*	Similar to DAS Less than 30 seconds	Approximately same as text based password	Hard to remember (to some extent), Dictionary attack, Very weak against

				video recording and spyware
TwoStep [123]	Simple Shoulder Surfing Resistance (to some extent)*	- Similar to Dhamija scheme For 1-2 rounds, it is approximately 30 seconds	$\binom{N}{K}^{\wedge r}$ Where: choosing K pass-images from a set of N images, and r is number of rounds	Guessing attack, Brute force search, Weak against shoulder surfing and spyware
GPI Graphical Password with Icons [4]	Similar to Pass-Point, but reduces Hot-spot attacks	17 seconds Average time to login	P(n, k) Permutation $\binom{n!}{(n-k)!}$ (Select in a specific order k pass-icons from n icons)	Brute force search, Shoulder surfing attack
PassShapes [131]	Fast to perform*	6.5 seconds Average time to login	N strokes, each stroke with 7 options. N^7	Guessing attack, Brute force search, Shoulder surfing attack

Note:
Shoulder Surfing Resistance (to some extent)*: provides shoulder surfing resistance for at least one attempt. However, the scheme does not prove to provide resistance for many video and spyware recording login sessions.
Fast to perform*: Takes approximately less than 30 seconds to perform.
Long time to login*: Takes approximately more than 30 seconds to perform

The following table compares graphical password schemes in terms of usability and security factors.

Table 4: Comparison of the text-based, 'Challenge questions' and graphical password schemes. The✓ means the scheme satisfies the criteria/feature, and × means the scheme does not satisfy the criteria/feature.

GP Scheme	Easy to do	Fast to perform *	(Easy to remember) Simple to, recall or recognition of a few pass-images*	Does not use extra devices	Large Password Size*	Provides resistance against key-logging	Provides resistance against one time shoulder surfing	Provides full resistance against spyware and many video recorded login sessions
Text-based	✓	✓	✓**	✓	✓	×	×	×
'Challenge Questions'	✓	✓	✓**	✓	✓***	×	×	×
Convex Hull Scheme [109]	✓	×	×	✓	✓	✓	✓	✓*
Draw a Pass (Haptic) [70]	✓	✓	✓	×	✓	✓	✓	✓*
Man [71] & Hong [49]	✓	✓	×	✓	✓	✓	✓	✓*
Association Based [66]	✓	✓	✓	✓	✓	✓	✓	×
Blonder [7]	✓	✓	✓	✓	✓	✓	×	×
Wiedenbeck et al. [133]	✓	×	×	✓	✓	✓	✓	✓*
Dhamija [27]	✓	✓	✓	✓	×	✓	×	×
Passfaces [88]	✓	✓	×	✓	×	✓	×	×
Pass-point [132]	✓	✓	✓	✓	✓	✓	×	×
Passlogix [89]	✓	✓	✓	✓	×	✓	×	×
S3PAS [139]	×	×	✓	✓	✓	✓	✓	✓*
DAS [58] (Jermyn)	✓	✓	✓	✓	✓	✓	×	×
Thorpe [119]	✓	✓	✓	✓	✓	✓	×	×
Recall-based Decoy Strokes [138]	✓	✓	✓	✓	✓	✓	✓	×
PassGo [115]	✓	✓	✓	✓	✓	✓	×	×
Recall Disappearing	✓	✓	✓	✓	✓	✓	✓	×

Stokes [138]								
Recall Line Snaking [138]	✓	✓	✓	✓	✓	✓	✓	✖
Syukri [114]	✓	✓	✓	✓	✓	✓	✖	✖
3D [1]	✓	✓	✓	✓	✓	✓	✖	✖
'Picture Password' [106]	✓	✓	✓	✓	✓	✓	✖	✖
Cognitive Auth Scheme [130]	✖	✖	✖	✓	✓	✓	✓	✖
NPSRSS [42]	✓	✓	✖	✓	✖	✓	✓	✖
DAGPS [41]	✓	✓	✓	✓	✓	✓	✓	✖
TwoStep [123]	✓	✓	✖	✓	✖	✓	✓	✖
GPI Graphical Password with Icons [4]	✓	✓	✖	✓	✓	✓	✖	✖
PassShapes [131]	✓	✓	✓	✓	✖	✓	✖	✖

Note:
Easy to do*: Takes 4 or less instructional steps to perform. The same instructional steps can be repeated if necessary.
Fast to Perform*: Takes approximately less than 30 seconds to perform.
Simple recall or recognition of a few pass-images*: Recall and/or recognize less than 5 objects
Large password size*: (Theoretical) At least 10^7 possibilities
✓*: provides shoulder surfing resistance for at least one attempt. However, the scheme does not prove to provide resistance for many video and spyware recorded login sessions.
✓**: The password can get complicated easily, or be forgotten. For the 'challenge-questions' method, it can be hard for the user to remember the exact answers to the questions.
✓***: The theoretical password space is close to the text-based password. However the 'challenge questions' answers are often related to the user's preference, favourites, items, objects, family, and personal information. That makes it very small password space and can be guessed if the attacker has more knowledge of the user.

The graphical password schemes address many security and usability issues. Most of the schemes addressed brute force search and guessing attacks but still remained vulnerable to predictable image selection and 'hot-spots'. Many schemes are weak against observation attacks including (i) shoulder surfing, (ii) eavesdropping and (iii) spyware. In terms of shoulder surfing resistance, most of the schemes addressed the issue. However, the schemes resist shoulder surfing for only a few login attempts, and

the majority were weak against captured login sessions. For spyware, all of the schemes indirectly give hints about the password and they are weak at addressing this issue. If an attacker observes or records the user's input and screen contents during many successful logins, information about the password can be acquired. From analyzing many graphical password schemes, and most of the recognition-based graphical password schemes and cognitive authentication scheme [130], we conclude that currently there is no graphical password scheme that (i) is easy to do without challenging steps, (ii) is fast to perform, (iii) does not use external or haptic devices, (iv) requires only simple recall or recognition of a few pass-images, (v) has a large password size, (vi) provides one-time shoulder surfing resistance, and (vii) provides full resistance against spyware and many video recorded login sessions. Therefore, it is still an open research area to find a graphical password scheme that satisfies the seven criteria mentioned above. In some ways, the field of user passwords is similar to criminology. In particular, during a crime, a criminal can indirectly leave traces of the crime committed and this small trace can then provide useful hints and information to determine the identity of criminal. The same goes for graphical passwords: the user's every movement can directly or indirectly provide hints to the observer about the password and pass-images.

5.5 Conclusion

It is a challenge to produce an online system that is easy to use, inexpensive, easy to deploy, and completely secure and trusted. However, the proposed authentication method aims to be easy to use, inexpensive, and easy to deploy, while enhancing security in online environments. The proposed method takes advantage of what is provided without adding additional devices, and equipment. Moreover, the video banking authentication method provides a transparent multimodal biometric continuous authentication system. In order to have a 'steps-free' user-friendly system, the fusion of voice, and face biometrics are used. To authenticate a user, an HD camera and a minimal user interaction, during a video banking conference is required. The method relies on the existing simple infrastructure of video conferencing and does not

require additional devices. The results of the fusion of the proposed multimodal biometrics indicate that the fused multimodal biometric system has a higher performance and accuracy rates than that of the unimodal biometrics. The proposed method aims to give a service provider more automatic authentication techniques for the authentication of online users with minimal user interaction. Also, the proposed method can function as an add-on to existing online transaction techniques; as an additional security layer it can minimize and resist online attacks, malware, and MITB attack. It is feasible to have a 'steps-free' multimodal biometric system based on physiological and behavioral biometrics. The proposed method provides an approach that effectively resists MITB attack and identifies users during a video banking conference with minimal user interaction.

To have a video banking session, the user must be logged in to online banking. Banks [92] often use the 'challenge questions' method in addition to text passwords. The 'challenge questions' method requires the user to answer the questions in the same way that they were answered during registration. The drawbacks of this method are that the answers can be forgotten and can be vulnerable to key-logging attacks. The answers are often related to the user and it is weak against shoulder-surfing attacks. Graphical password schemes can address some drawbacks of the 'challenge questions' method, such as key-logging attacks. However, none of the analyzed schemes address all security and usability factors. The choice of the graphical password scheme depends on the usability and security requirements set by the online authentication system. For the video banking system, the user must be logged in to online banking in order to initiate a video banking session. We recommend combining a text-based password with a graphical password. By having the user login to online banking by using a combination of text-password and a graphical password, we are having a large password size and preventing key-logging attacks [123]. During the video banking session, the bank identifies a user based on the user's voice, and face biometrics. These combined biometrics form a transparent multimodal biometric system that provides continuous authentication and addresses MITB attacks. Combining the authentication technique to login to online banking with the transparent authentication technique during the video banking session provides a multimodal biometric multi-factor authentication system.

Chapter 6: Conclusion and Future Work

In this chapter we provide the conclusion, lessons learned, and discuss several open problems related to the work of the thesis, and future work.

6.1 Conclusion

Without a doubt, the internet has changed users' lifestyles, businesses, organizations, and governments. However, its benefits come with drawbacks, such as incomplete security and trust. The proposed TMBA system provides a transparent multimodal biometric continuous authentication system. It takes advantage of what is provided without requiring additional devices. To authenticate a user, an HD camera during a video banking conference, is required. The system fuses the voice and face biometrics. The results of the fusion of the multimodal biometrics indicate that the fused multimodal biometric system has a higher performance and accuracy rates than that of the unimodal biometrics. The proposed system provides an approach that identifies users during a video conference with minimal user interaction.

The proposed end-to-end trust establishment procedure enables users that have never met before to verify each other's identity during a video conference and enhance their trust in one another. After first time authentication, the system does not require any user cooperation; the user does not have to repeat steps in order to verify the identity claim of the other party as this will be done automatically by the system. Furthermore, the trust establishment procedure is easily deployable since it uses the current simple infrastructure of a video conferencing system. The system is practical, since the solution does not depend on any specific third party. It can be used as an additional security layer to authenticate the identity at the other end. To collect information about a user, the proposed system depends on multiple third-parties; however, it does not depend on a specific one. If a third-party is down, the system still works and relies on the available

third-parties. Without fully depending on a specific third-party, it is a challenge to have a complete end-to-end security and trust system for users that are unknown to each other. This area of research is a security and trust challenge that still needs to be addressed. However, the proposed end-to-end identification and trust establishment procedure provides an approach to authenticate unknown users to each other during a 'user-to-user' video conference.

The proposed 'user-to-service' video banking authentication method takes advantage of what is provided without requiring additional devices and equipment. Moreover, the video banking authentication method provides a transparent multimodal biometric continuous authentication system. To authenticate a user, an HD camera is required. The method relies on the existing simple infrastructure of video conferencing and does not require additional devices. The method aims to give a service provider more automatic authentication techniques for the authentication of online users with minimal user interaction. Also, the proposed method can function as an add-on to existing online transaction techniques; as an additional security layer it can minimize and resist online attacks, malware, and MITB attacks. It is feasible to have a 'steps-free' multimodal biometric system based on physiological and behavioral biometrics. The proposed method provides an approach that effectively resists MITB attacks and identifies users during a video banking conference with minimal user interaction.

To have a video banking session, the user must be logged in to online banking. Banks [92] often use the 'challenge questions' method in addition to text passwords. The 'challenge questions' method requires the user to answer the questions in the same way that they were answered during registration. The drawbacks of this method are that the answers can be forgotten and can be vulnerable to key-logging attacks. The answers are often related to the user and it is weak against shoulder-surfing attacks.

Graphical password schemes can address some drawbacks of the 'challenge questions' method, such as key-logging attacks. However, they have their own drawbacks. From analyzing many graphical password schemes, and cognitive authentication scheme [130], we conclude that currently there is no graphical password scheme that (i) is easy to do without challenging steps, (ii) is fast to perform, (iii) does

not use external or haptic devices, (iv) requires only simple recall or recognition of a few pass-images, (v) has a large password size, (vi) provides one-time shoulder surfing resistance, and (vii) provides full resistance against spyware and many video recorded login sessions. Therefore, it is still an open research area to find a graphical password scheme that satisfies the seven criteria mentioned above.

For the video banking system, the user must be logged in to online banking in order to initiate a video banking session. We recommend using text-based password with a graphical password. The text-based password can provide a large password size, and a graphical password scheme can prevent key-logging attacks and could address phishing and shoulder-surfing attacks [123]. The choice of the graphical password scheme depends on the usability and security requirements set by the online authentication system.

During the video banking session, the bank identifies a user based on the user's voice and face biometrics. These combined biometrics form a transparent multimodal biometric system that provides continuous authentication. Combining the authentication technique (text-based password with a graphical password) to login to online banking with TMBA technique during the video banking session provides a multimodal biometric multi-factor authentication system. The proposed system provides an approach that minimizes online attacks, malware, and effectively resists key-logging attacks and MITB attack, and identifies users during a video banking conference with minimal user interaction.

Our main contributions, in summary, are (i) a TMBA system that transparently authenticates a user during a video conference, (ii) an end-to-end trust establishment procedure to enable users that have never met before to verify each other's identity during a video conference and enhance their trust in one another, (iii) a video banking authentication method, (iv) an analysis of various graphical password schemes and the conclusion that none satisfies all the usability and security factors; and (v) a multimodal biometric multi-factor continuous authentication system.

Designing a completely secure and trusted system is a challenge that still needs to be addressed; various enhancements are necessary to provide more security and trust. To enhance authentication in online environments, we propose authentication techniques for various applications, including 'user-to-user' and 'user-to- service' systems. The proposed authentication techniques aim to enhance security and trust for video applications in the untrusted online environment. The proposed systems aim to address: (i) compromised hosts, the MITB attack, and web attacks, (ii) usability in online authentication, and (iii) enhancing trust in online environments.

In the security world, there are many open problems in the real-world that required to be fully addressed. The open problems include: (i) compromised hosts, (ii) malware and web attacks, (iii) spyware, (iv) semantic attacks (eg. phishing attack), (v) system error or vulnerability to lead to potential attacks, (vi) usability, (vii) naïve users, and (viii) digital trust [14, 72]. Each of them is an open problem leading to insecure and untrusted online systems in the real-world. Many proposed academic and commercial systems focus on one or few open problems only, without addressing other open problems. For example, many proposed graphical password schemes address security concerns without addressing usability concerns. On the other hand, many address usability concerns without addressing security concerns. Furthermore, those that address usability and security concerns do not address all security concerns, thus proposing incomplete systems. Realistic assumptions must be a main part of designing secure and trusted online systems. Academic proposals must have more realistic assumptions that are reasonable to the service provider and the user. For example, to address MITB attacks, many academic proposals provide a complete solution; however they assume the user to have a malware-free mobile. Such academic proposals are acceptable in the academic world; however they may not be possible in the real-world.

In our proposals, we do not claim to provide completely secure and trusted systems, nor do we claim our assumptions are completely reasonable in real-world. However, we aimed to provide enhancements to have secure and trusted online systems, and aimed to make our assumptions as acceptable as much as we can. Designing a completely secure and trusted system is a challenge that still needs to be addressed.

However to have such a system, we need to address all security, trust, and privacy open problems. Once we provide solutions to each of the security, trust, and privacy open problems, then we can have a very secure and trusted system. We also have to make system assumptions very realistic to be adapted by the real-world. Without having the open problems solved, it is still a challenge to have a completely secure and trusted system. However, we still require enhancements to provide acceptable and more secure online system.

6.2 Future Work

Currently, each authentication technique including multi-factor authentication, has its benefits and drawbacks. There is no complete security solution that: (i) is easy to use, (ii) is easy to deploy, (iii) has a low cost and (iv) prevents all types of security holes, attacks and vulnerabilities. Moreover, there is no authentication system that is suitable for all types of online applications. Each system has to take advantage of what is provided in order to authenticate a user, without having to add devices and requirements. For example, for video conferencing applications, an authentication system can take advantage of the user's biometrics, such as face, ear, and voice biometrics, which are presented in the media. For TMBA system, we will focus on using better recognition methods to enhance performance by increasing GAR and reducing FAR. Enhancing the detection and recognition of face and voice biometrics will boost the performance of the system; the aim is to make FAR < 0.1% and for GAR to reach 100%. For face recognition, we will use the Fisherface method instead of EigenFace. Moreover, we can add more available biometrics, such as, ear biometrics, shown in Appendix C; this will be in order to fuse it with face and voice biometrics to provide a transparent multimodal biometric system that is based on three human biometrical traits. For the end-to-end trust establishment procedure for a transparent multimodal biometric system, the focus will be on implementing a complete system that includes reputation systems and end-to-end secure video calls. The proposed system can use multiple third-parties; however, it does not depend on a specific one. Without fully

depending on a specific third-party, it is a challenge to have a complete end-to-end security and trust system for users that are unknown to each other. This area of research is a security and trust challenge that still needs to be addressed.

For the proposed video banking authentication method, a complete implementation of end-to-end video security between two ends is required. We aim to have optimal security without requiring additional steps for the user to keep the method user friendly. For graphical password schemes, further research is required to enhance usability and security. There are few shoulder surfing resistant graphical password schemes, and further work is required in this field. The design of graphical password schemes that address all usability and security concerns is a research area that still needs to be addressed.

References

[1] F.A. Alsulaiman, A. El Saddik, "Three-Dimensional Password for More Secure Authentication," IEEE Transactions on Instrumentation and Measurement, vol.57, no.9, pp.1929-1938, Sept. 2008

[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.19, no.7, pp.711-720, Jul 1997

[3] F. Besbes, H. Trichili, B. Solaiman, "Multimodal Biometric System Based on Fingerprint Identification and Iris Recognition," 3rd International Conference on Information and Communication Technologies: From Theory to Applications, Damascus, 2008. ICTTA 2008, pp.1-5, 7-11 April 2008.

[4] K. Bicakci, N.B. Atalay, Mustafa Yuceel, H. Gurbaslar, B. Erdeniz, "Towards Usable Solutions to Graphical Password Hotspot Problem," 33rd Annual IEEE International Computer Software and Applications Conference, 2009. COMPSAC '09. vol.2, pp.318-323, 20-24 July 2009

[5] A. Bhargav-Spantzel, A. C. Squicciarini and E. Bertino, "Trust Negotiation in Identity Management," IEEE Security & Privacy, vol. 5, no. 2, pp. 55-63, March-April 2007.

[6] R. Biddle, S. Chiasson, P.C. van Oorschot. "Graphical Passwords: Learning from the First Twelve Years," ACM Computing Surveys 44(4), 2012 (to appear, 2012).

[7] G. E. Blonder, "Graphical passwords," in Lucent Technologies, Inc., Murray Hill, NJ, U. S. Patent, Ed. United States, 1996.

[8] B. Bradway, "Transforming Branch Banking: Real-Time Advice in the 21st Century," TANDBERG, August 2005. [Online]. Available: http://www.banktech.com/whitepaper/Architecture/Infrastructure/transforming-branch-banking:-real-time-advice-in-wp9600003. [Accessed 07 October 2011].

[9] A. Brooks, L. Gao. "Face Recognition". [Online], Available: http://dailyburrito.com/projects/FaceRecognition.pdf [Accessed 4 May 2011]

[10] S. Cadavid, M.H. Mahoor, M. Abdel-Mottaleb, "Multi-modal Biometric Modeling and Recognition of the Human Face and Ear," IEEE International Workshop on Safety, Security & Rescue Robotics (SSRR), 2009, pp.1-6, 3-6 Nov. 2009

[11] E. Chandra, C. Sunitha, "A review on Speech and Speaker Authentication System using VoiceSignal feature selection and extraction," IEEE International Advance Computing Conference, 2009. IACC 2009, pp.1341-1346, 6-7 March 2009

[12] K. Chang, K.W. Bowyer, S. Sarkar, B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.25, no.9, pp. 1160- 1165, Sept. 2003

[13] S. Charney, "Establishing End to End Trust," Microsoft Corp., 2008. [Online]. Available: download.microsoft.com/download/7/2/3/723a663c-652a-47ef-a2f5-91842417cab6/Establishing_End_to_End_Trust.pdf. [Accessed 03 January 2011].

[14] S. Charney, "Creating a More Trusted Internet," Microsoft Corp., 8 April 2008. [Online]. Available: download.microsoft.com/download/2/f/7/2f752ae4-7e1d-4dbd-b75a-aa2dcb0eff5b/End_to_End_Trust_Statement_of_Purpose_Charney.pdf. [Accessed 07 April 2011].

[15] S. Charney, "Rethinking the Cyber Threat: A Framework and Path Forward," Microsoft Corporation, 2009. [Online]. Available: download.microsoft.com/download/F/1/3/F139E667-8922-48C0-8F6A-B3632FF86CFA/rethinking-cyber-threat.pdf. [Accessed 08 March 2011].

[16] "Chat face to face with family and friends," [Online]. Available: http://www.google.com/chat/video [Accessed 10 11 2011].

[17] S. Chaudhary, R. Nath, “A Multimodal Biometric Recognition System based on Fusion of Palmprint, Fingerprint and Face,” International Conference on Advances in Recent Technologies in Communication and Computing, 2009, ARTCom '09, pp.596-600, 27-28 Oct. 2009.

[18] D.S. Chen; Z.K. Liu, "Generalized Haar-Like Features for Fast Face Detection," International Conference on Machine Learning and Cybernetics, 2007, vol.4, no., pp.2131-2135, 19-22 Aug. 2007

[19] H.T Cheng, Y.H. Chao, S.L. Yeh; C.S. Chen, H.M Wang, Y.P. Hung, "An Efficient Approach to Multimodal Person Identity Verification by Fusing Face and Voice Information," IEEE International Conference on Multimedia and Expo, 2005. ICME 2005. pp.542-545, 6-6 July 2005

[20] Cisco’s Technology News Site. “Citizens Financial Group launches video banking pilot in four states”. [Online], Available: http://newsroom.cisco.com/press-release-content?type=webcontent&articleId=661350 [Accessed 2 November 2011]

[21] H.P. Combrinck, E.C. Botha, “On The Mel-scaled Cepstrum,” Proceedings of the Seventh Annual South African Workshop on Pattern Recognition, University of Pretoria, Pretoria.

[22] N. Covavisaruch, P. Prateepamornkul, "Personal Identification System using Hand Geometry and Iris Pattern Fusion," IEEE International Conference on Electro/information Technology, 2006, pp.597-602, 7-10 May 2006

[23] N. Covavisaruch, P. Prateepamornkul, P. Ruchikachorn, P. Taksaphan, “Personal Verification and Identification Using Hand Geometry,” Transaction on Computer and Information Technology. ECTI, 2005, vol.1, no.2, pp. 134-140, November 2005

[24] N. Covavisaruch, C. Saengpanit, “Time Stamp Detection and Recognition in Video Frames,” Proceedings, International Conference on Imaging Science, Systems, and Technology. CISST’04, Las Vegas, Nevada, pp. 173-178, 21-24 June 2004.

[25] J.G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.15, no.11, pp.1148-1161, Nov 1993

[26] D. Davis, F. Monrose, M. K. Reiter, "On user choice in graphical password schemes," Proceedings of the 13th USENIX Security Symposium, San Diego, CA, USA, pp. 151–164, August 2004

[27] R. Dhamija, A. Perrig, "Deja Vu: A User Study Using Images for Authentication," Proceeding SSYM'00 Proceedings of the 9th conference on USENIX Security Symposium - Volume 9 USENIX Association Berkeley, CA, USA, 2000

[28] M.N. Doja, N. Kumar, "Image Authentication Schemes against Key-Logger Spyware," Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2008. SNPD '08. pp.574-579, 6-8 Aug. 2008

[29] "Entrust IdentityGuard," [Online]. Available: http://www.entrust.com/strong-authentication/identityguard/tokens/. [Accessed 11 12 2011].

[30] Entrust Inc. "Stong Authentication". [Online]. Available: http://www.entrust.com/authentication/index.htm. [Accessed 05 02 2011].

[31] Entrust Inc. "Strong Authentication Methods". [Online]. Available: http://www.entrust.com/strong-authentication/authenticators.htm. [Accessed 10 09 2010].

[32] C.E. Erdem, S. Ulukaya, A. Karaali, A.T. Erdem, "Combining Haar Feature and skin color based classifiers for face detection," International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE, pp.1497-1500, 22-27 May 2011

[33] M.N. Eshwarappa, M.V. Latte, "Bimodal Biometric Person Authentication System Using Speech and Signature Features," International Journal of Biometrics and Bioinformatics, (IJBB), Volume (4): Issue (4), 147-160, 2010

[34] K. M. Everitt, T. Bragin, J. Fogarty, T. Kohno, "A Comprehensive Study of Frequency, Interference, and Training of Multiple Graphical Passwords," In Proceedings of the 27th SIGCHI International Conference on Human Factors in Computing Systems (2009), 889-898.

[35] "Eye detection", [Online], Available: http://yushiqi.cn/research/eyedetection [Accessed 08 09 2011].

[36] "Face Detection", [Online], Available: http://www.cvip.uofl.edu/wwwcvip/education/ECE523/Spring%202011/Lec7.pdf [Accessed 06 October 2011].

[37] S. Farmand, O. Bin Zakaria, "Improving graphical password resistant to shoulder-surfing using 4-way recognition-based sequence reproduction (RBSR4)," The 2nd IEEE International Conference on Information Management and Engineering (ICIME), 2010, pp.644-650, 16-18 April 2010

[38] "FirstOntario Credit Union Introduces Video Banking". [Online], Available: http://banknerd.ca/2010/07/16/firstontario-credit-union-introduces-video-banking/ [Accessed 22 November 2011].

[39] N.A. Fox, R. Gross, J.F. Cohn, R.B. Reilly, "Robust Biometric Person Identification Using Automatic Classifier Fusion of Speech, Mouth, and Face Experts," IEEE Transactions on Multimedia, vol.9, no.4, pp.701-714, June 2007

[40] R.W. Frischholz, U. Dieckmann, "BiolD: a multimodal biometric identification system," Computer, vol.33, no.2, pp.64-68, Feb 2000

[41] H. Gao, X. Liu, S. Wang, H. Liu, R. Dai , "Design and Analysis of a Graphical Password Scheme," Fourth International Conference on Innovative Computing, Information and Control (ICICIC), 2009, pp.675-678, 7-9 Dec. 2009

[42] H. Gao; Z. Ren; X. Chang; X. Liu; Aickelin, U., "A New Graphical Password Scheme Resistant to Shoulder-Surfing," International Conference on Cyberworlds (CW), 2010, pp.194-199, 20-22 Oct. 2010

[43] "Getting started with 2-step verification," [Online]. Available:http://support.google.com/accounts/bin/static.py?hl=en&page=guide.cs&guide=1056283. [Accessed 11 12 2011].

[44] M.K Gill, R. Kaur, J. Kaur, "Vector Quantization based Speaker identification", International Journal of Computer Applications, vol.4, no.2, article 1, pp. 1-4, July 2010

[45] P. Golle; D. Wagner, "Cryptanalysis of a Cognitive Authentication Scheme (Extended Abstract)," IEEE Symposium on Security and Privacy, 2007, SP '07, pp.66-70, 20-23 May 2007

[46] K. Haseeb, M. Arshad, S. Ali, S. Yasin. "Secure E-commerce Protocol," International Journal of Computer Science and Security, vol.5, no.1, pp.742-751, April 2011

[47] E. Hjelmas, B.K. Low, "Face Detection: A Survey", Computer Vision and Image Understanding, Vol. 83, No. 3, pp.236-274, 2001

[48] L. Hong, A. Jain, "Integrating faces and fingerprints for personal identification," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.20, no.12, pp.1295-1307, Dec 1998

[49] D. Hong, S. Man, B. Hayes, M. Matthews, "A password scheme strongly resistant to spyware", Proc. Int. Conf. on Security and Management, Las Vegas, 2004, pp. 94-100

[50] R.L. Hsu, M. Abdel-Mottaleb, A.K. Jain, "Face detection in color images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, no.5, pp.696-706, May 2002

[51] IBM Internet Security Systems. "Beating the Man-in-the-browser with a ZTIC," [Online], Available: http://blogs.iss.net/archive/ZTIC.html [Accessed 6 April 2012]

[52] Internet2.edu. "A Framework of Requirements, Threat Models, and Security Services for Videoconferencing over Internet2," [Online], Available: http://middleware.internet2.edu/video/draftdocs/draft-chatterjee-johnson-vc-security-01.html [Accessed 14 March 2012]

[53] K. Irwin, T. Yu, "Preventing attribute information leakage in automated trust negotiation," Proceedings of the 12th ACM conference on Computer and communications security, pp. 36 - 45, 2005.

[54] A. Jain, K. Nandakumara, A. Ross, "Score normalization in multimodal biometric systems", The Journal of the Pattern Recognition Society, vol. 38, p.2270, 2005.

[55] A.K. Jain, A. Ross, S. Prabhakar, "Biometrics: A Tool for Information Security," IEEE Transactions on Information Forensics and Security, vol.1, no.2, pp.125-143, June 2006.

[56] K. Jain, A. Ross, S. Prabhakar, “Introduction to Biometric Recognition”, IEEE transactions on circuit and systems for video technology, Vol. 14, no. 1, January 2004.

[57] A. Javadtalab, L. Abbadi, M. Omidyeganeh, S. Shirmohammadi, C.M. Adams, A. El Saddik, "Transparent non-intrusive multimodal biometric system for video conference using the fusion of face and ear recognition," Ninth Annual International Conference on Privacy, Security and Trust (PST), 2011, pp.87-92, 19-21 July 2011.

[58] I. Jermyn, A. Mayer, F. Monrose, M.K. Reiter, A.D. Rubin, “The Design and Analysis of Graphical Passwords”, Proceedings of the 8th USENIX Security Symposium, Washington, D.C., USA, August 23–26, 1999.

[59] R. Kim, "Expert: iPhone vulnerable to text message attack," SFGate, 30 July 2012. [Online]. Available: http://blog.sfgate.com/techchron/2009/07/30/expert-iphone-vulnerable-to-text-message-attack/. [Accessed 14 January 2011].

[60] A. Klenk, G. Carle, B. Radier and M. Salaun, "Secure Statless Trust Negotiation," In IFIP Network and Service Security Conference, Paris, France, June 2009.

[61] T. Ko, “Multimodal Biometric Identification for Large User Population Using Fingerprint, Face, and Iris Recognition”, Proceedings of the 34th Applied Imagery and Pattern Recognition Workshop, pp. 218-223, 2005.

[62] K. M. Lavassani, B. Mohavedi and V. Kumar, "Identification in Electronic Networks: Characteristics of e-Identifiers," ICEC '06 Proceedings of the 8th international conference on Electronic commerce, ACM, New York, NY, pp. 216 - 224, 2006.

[63] H.J Lee; W.S Lee, J.H. Chung, "Face recognition using Fisherface algorithm and elastic graph matching," Proceedings International Conference on Image Processing, 2001, pp.998-1001, 2001

[64] A. J. Lee, M. Winslett, J. Basney, V. Welch, "The Traust Authorization Service," ACM Transactions on Information and System Security (TISSEC), vol.11, no.2, issue 1, pp. 1-33, Feb 2008.

[65] J. Li, N. Li and W. H. Winsborough, "Automated trust negotiation using cryptographic credentials," Proceedings of the 12th ACM conference on Computer and communications security, pp. 46 - 57, 2005.

[66] Z. Li, Q. Sun, Y. Lian, D.D. Giusto, "An Association-Based Graphical Password Design Resistant to Shoulder-Surfing Attack," IEEE International Conference on Multimedia and Expo, 2005. ICME 2005, pp. 245- 248, 6-8 July 2005

[67] Z. Li, L. Xue, F. Tan, "Face detection in complex background based on skin color features and improved AdaBoost algorithms," IEEE International Conference on

Progress in Informatics and Computing (PIC), 2010, vol.2, no., pp.723-727, 10-12 Dec. 2010

[68] R. Lienhart, A. Kuranov, V. Pisarevsky, “Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection, ” Proceeding of the 25th German Association for Pattern Recognition DAGM (Deutsche Arbeitsgemeinschaft für Mustererkennung DAGM e.V.), Magdeburg, Germany, 2003

[69] R. Lienhart, *J.* Maydt, "An extended set of Haar-like features for rapid object detection," International Conference on Image Processing, 2002, pp. I-900- I-903 vol.1, 2002

[70] B. Malek, M. Orozco, A. El Saddik, "Novel Shoulder-Surfing Resistant Haptic-based Graphical Password", Proceedings of the Eurohaptics 2006 Conference, Paris, France, July 2006

[71] S. Man, D. Hong, M. Matthews, “A shoulder-surfing resistant graphical password scheme - WIW”, Proc. Int. Conf. on Security and Management, Las Vegas, 2003, pp. 105-111

[72] M. Mannan, "Authentication and Securing Personal Information in an Untrusted Internet," Ph.D thesis, School of Computer Science, Carleton University, Ottawa, Canada, 2009.

[73] Marvis.com. “Man in the Browser Attack (MITB),” [Online], Available: http://www.maravis.com/library/man-in-the-browser-attack/ [Accessed 9 January 2012]

[74] X. Meng, “Study on the Model of E-Commerce Identity Authentication based on Multi-biometric Features Identification,” International Colloquium on Computing, Communication, Control, and Management, Guangzhou, pp. 196 – 200, 2008.

[75] M.M. Monwar, M.L. Gavrilova, “Multimodal Biometric System Using Rank-Level Fusion Approach,” IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol.39, Issue 4, pp. 867 – 878, 2009.

[76] E. Moser, “Amygdala activation at 3T in response to human and avatar facial expressions of emotions,” Journal of Neurosci Methods (2006)

[77] L. Muda, M. Begam, I. Elamvazuthi, “Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques”, Journal of Computing, vol.2, Issue 3, March 2010

[78] D. Nali, J. Thorpe, "Analyzing User Choice in Graphical Passwords," Tech. Report, School of Computer Science, Carleton University, Canada, 2004. [Online]. Available:

http://www.cs.carleton.ca/research/tech_reports/2004/TR-04-01.pdf. [Accessed 04 01 2011].

[79] E. Naone, ""SMS of Death" Could Crash Many Mobile Phones," Technology Review, 4 January 2011. [Online]. Available: http://www.technologyreview.com/printer_friendly_article.aspx?id=27021. [Accessed 07 January 2011].

[80] M. Nilsson, "Speaker Verification in JAVA," M.S. thesis, School of Microelectronic Engineering, Griffith University, Brisbane, Australia, 2001.

[81] G. Ollmann. “Man-in-the-Browser Attack Vectors”. [Online], Available: http://www.slideshare.net/guestb1956e/csi2008-gunter-ollmann-maninthebrowser-presentation [Accessed 9 April 2012]

[82] “Opencv”, [Online], Available: https://code.ros.org/svn/opencv/trunk/opencv/data/haarcascades/ [Accessed 11 10 2011].

[83] “OpenCV”, [Online], Available: http://opencv.willowgarage.com/wiki/ [Accessed 11 10 2011].

[84] OWASP.org. “Session hijacking attack,” [Online], Available: https://www.owasp.org/index.php/Session_hijacking_attack [Accessed 4 March 2011]

[85] F. Paci, D. Bauer, E. Bertino, D. M. Blough and A. Squicciarini, "Minimal credential disclosure in Trust Negotiations," DIM '08 Proceedings of the 4th ACM workshop on Digital identity management, pp. 89-96, 2008.

[86] P. Park. “Voice over IP Security”. [Online], Available: http://cdn.ttgtmedia.com/searchSecurity/downloads/1587054698_chapter_2_patrick_park.pdf [Accessed 19 May 2012]

[87] D. Parlanti, D. Guli and M. C. Pettenati, "Intermediation for trust-enabling networked decentralized exchange systems," ACM Proceedings of the 13th Eurpoean conference on Cognitive ergonomics, pp. 64-70, 2006.

[88] Passfaces. "Passfaces: Two Factor Authentication for the Enterprise,". [Online]. Available: http://www.realuser.com/. [Accessed 10 02 2011].

[89] "Passlogix," [Online]. Available: http://www.oracle.com/us/corporate/Acquisitions/passlogix/index.html. [Accessed 02 09 2011].

[90] S. Phung, D. Chai, A. Bouzerdoum, "Skin colour based face detection," Intelligent Information Systems Conference, The Seventh Australian and New Zealand 2001, pp. 171- 176, 18-21 Nov. 2001

[91] N. Raleigh, "Coastal's new teller system increases branch traffic," Credit Union National Association, 5 November 2009. [Online]. Available: http://www.cuna.org/newsnow/09/system050809-6.html. [Accessed 11 January 2011].

[92] RBC Royal Bank. "Online Banking". [Online]. Available: http://www.rbcroyalbank.com/online/online-banking-security-guarantee.html. [Accessed 11 02 2011].

[93] "Requirements, Threat Models, and Security Services for Videoconferencing over Internet2", [Online], Available: http://middleware.internet2.edu/video/draftdocs/draft-chatterjeejohnson-vc-security-01.html [Accessed 03 05 2012]

[94] "Revolutionizing the Banking World," uGenius, [Online]. Available: http://www.ugenius.com/products_ptm_description2.php. [Accessed 30 January 2012].

[95] D. A. Reynolds and L. P. Heck, "Automatic Speaker Recognition: Recent Progress, Current Applications, and Future Trends," 19 February 2001.

[96] A. Ross, R. Govindarajan, "Feature Level Fusion Using Hand and Face Biometrics", proc. Of SPIE Conference on Biometric Technology for Human Identification II, Vol. 5779, pp. 196-204, Orlando, USA, 2005

[97] A. Ross, A. Jain, "Information fusion in biometrics", Pattern Recognition Letters, pp. 2115-2125, 2003. 2003

[98] T. Ryutov, L. Zhou, C. Neuman, T. Leithead, K. E. Seamons, "Adaptive Trust Negotiation and Access Control," Proc. Of the 10th ACM Symposium on Access Models and Technologies, Stickholm, Sweden, June 2005.

[99] M.C. Santana, O. D´eniz-Su´arez, L. Ant´on-Canal´ıs , J. Lorenzo-Navarro, "Face and Facial Feature Detection Evaluation," III International Conference on Computer Vision Theory and Applications. VISAPP'2008, Funchal, Portugal, January 2008.

[100] C. Santana, M. L. Navarro, J.H. Sosa, "An Study on Ear Detection and its Applications to Face Detection," Proceedings of the 14th international conference on Advances in artificial intelligence: spanish association for artificial intelligence, CAEPIA'11, La Laguna, Spain, pp. 313-322, November 2011.

[101] K. S Haseeb, M. Arshad, S. Ali, S. Yasin, "Secure E-Commerce Protocol," International Journal of Computer Science and Security, vol.5, no.1, pp. 742-751, April 2011.

[102] "Security," [Online]. Available: http://www.via3.com/info/products/features/security.aspx [Accessed 09 11 2011].

[103] G. Shangfu, L. Jun, S. Yizhen, "Design and Implementation of Anti-Screenshot Virtual Keyboard Applied in Online Banking," International Conference on E-Business and E-Government (ICEE), 2010, pp.1320-1322, 7-9 May 2010

[104] M. Sharkas, M.A. Elenien, "Eigenfaces vs. fisherfaces vs. ICA for face recognition; a comparative study," 9th International Conference on Signal Processing, 2008. ICSP 2008 , pp.914-919, 26-29 Oct. 2008

[105] R. N. Shepard, "Recognition memory for words, sentences, and pictures," Journal of Verbal Learning and Verbal Behavior, vol. 6, pp. 156-163, 1967.

[106] S. Sinofsky, "Signing in with a picture password," Microsoft, 2011. [Online]. Available: http://blogs.msdn.com/b/b8/archive/2011/12/16/signing-in-with-a-picture-password.aspx. [Accessed 25 12 2011].

[107] "Skype Security," [Online]. Available: http://www.skype.com/intl/en-us/security/ [Accessed 0807 2011].

[108] "Skype Security: Protecting your online safety, security and privacy," Skype, 2012. [Online]. Available: http://www.skype.com/intl/en-us/security/ . [Accessed 03 January 2012].

[109] L. Sobrado and J.C. Birget, "Graphical passwords," The Rutgers Scholar, An Electronic Bulletin for Undergraduate Research, vol. 4, 2002.

[110] "Speaker Recognition," Biometrics, 7 August 2006. [Online]. Available: http://www.biometrics.gov/Documents/speakerrec.pdf. [Accessed 09 April 2011].

[111] Spoken Language Systems. “Feature Extraction from Speech”, [Online], Available: http://www.lsv.uni-saarland.de/dsp_ss05_chap9.pdf [Accessed 12 November 2011].

[112] X. Suo; Y. Zhu; G.S. Owen, "Graphical passwords: a survey," 21st Annual Computer Security Applications Conference, pp.10 pp.-472, 5-9 Dec. 2005

[113] D.L. Swets, J.J. Weng, "Using discriminant eigenfeatures for image retrieval," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.18, no.8, pp.831-836, Aug 1996

[114] A.F. Syukri, E. Okamoto, M. Mambo, "A User Identification System Using Signature Written with Mouse," Proceeding ACISP '98 Proceedings of the Third Australasian Conference on Information Security and Privacy, Springer-Verlag London, UK, 1998

[115] H. Tao, "Pass-Go, a New Graphical Password Scheme," M.ASc. thesis, School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada, 2006.

[116] Techrepublic.com. “Set up secure anywhere video conferencing,” [Online], Available: http://www.techrepublic.com/article/set-up-secure-anywhere-video-conferencing/6083185 [Accessed 16 January 2012]

[117] "TelePresence: In-Person Experiences for All," [Online]. Available: http://www.cisco.com/web/telepresence/video-conferencing.html [Accessed 12 12 2011].

[118] J. Thorpe, P.C. van Oorschot, “Human-Seeded Attacks and Exploiting Hot-Spots in Graphical Passwords”, 16th USENIX Security Symposium, August 2007

[119] J. Thorpe, P.C. van Oorschot, "Towards secure design choices for implementing graphical passwords," 20th Annual Computer Security Applications Conference, 2004, pp. 50- 60, 6-10 Dec. 2004

[120] M. Turk, A. Pentland, “Eigenfaces for recognition”, Journal of Cognitive Neuroscience archive, vol.3, issue 1, Winter 1991, MIT Press Cambridge, MA, USA

[121] M. Turk, A. Pentland, "Face recognition using eigenfaces," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991. CVPR '91, pp.586-591, 3-6 Jun 1991

[122] Ubc.ca. "As new platforms of communication come online, it’s important to know the risks that may be involved". [Online], Available:http://www.publicaffairs.ubc.ca/2011/12/20/people-more-likely-to-lie-when-texting-ubc-research/[Accessed 5 January 2011].

[123] P.C. van Oorschot, T. Wan. “TwoStep: An Authentication Method Combining Text and Graphical Passwords,” MCETECH 2009: 4th International MCETECH Conference on eTechnologies. May 2009.

[124] "Video calling on Skype," Skype, 2012. [Online]. Available: http://www.skype.com/intl/en-us/features/allfeatures/video-call/. [Accessed 03 January 2012].

[125] "Video Calls," [Online]. Available: http://jitsi.org/ [Accessed 12 11 2011].

[126] "Video Conferencing," [Online]. Available: http://www.nefsis.com/ [Accessed 09 10 2011].

[127] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001, pp. I-511- I-518 vol.1, 2001

[128] "Voice over IP security", [Online], Available: http://cdn.ttgtmedia.com/searchSecurity/downloads/1587054698_chapter_2_patrick_park.pdf [Accessed 02 06 2012]

[129] Y. Wang, T. Tan, A.K. Jain, "Combining face and iris biometrics for identity verification", AVBPA'03 Proceedings of the 4th international conference on Audio- and video-based biometric person authentication, Springer-Verlag, pp. 805-813, Berlin, 2003.

[130] D. Weinshall, "Cognitive authentication schemes safe against spyware," IEEE Symposium on Security and Privacy, 2006, pp.6 pp.-300, 21-24 May 2006

[131] R. Weiss, A.D. Luca, "PassShapes - Utilizing Stroke Based Authentication to Increase Password Memorability," ACM NordiCHI 2008: Using Bridges, Lund, Sweden, pp. 383-392, 18-22 October.

[132] S. Wiedenbeck, J. Waters, J.C. Birget, A. Brodskiy, N. Memon, "Authentication using graphical passwords: Basic results", Human-Computer Interaction International, 2005, Las Vegas, July 25-27 2005

[133] S.Wiedenbeck, J. Waters, L. Sobrado and J.C. Birget, "Design and Evaluation of a Shoulder-Surfing Resistant Graphical Password Scheme," ACM Proceedings of the working conference on Advanced visual interfaces. AVI '06, pp. 177-184, 2006

[134] X. Xu; Z. Mu, "Feature Fusion Method Based on KCCA for Ear and Profile Face Based Multimodal Recognition," IEEE International Conference on Automation and Logistics, 2007, pp.620-623, 18-21 Aug. 2007

[135] X. Xu , Z. Mu, "Multimodal Recognition Based on Fusion of Ear and Profile Face," Proceedings of the Fourth International Conference on Image and Graphics, pp.598-603, August 22-24 2007

[136] M. H. Yang, D. Kriegman, N. Ahuja, "Detecting Faces in Images: A Survey", IEEE Trans. PAMI, vol.24, no.1, pp.34-58, Jan. 2002.

[137] L. Yuan, Z. Mu, X. Xu, "Multimodal recognition based on face and ear," International Conference on Wavelet Analysis and Pattern Recognition, 2007. ICWAPR '07, vol.3, pp.1203-1207, 2-4 Nov. 2007

[138] N.H Zakaria, D.Griffiths, S. Brostoff, J. Yan, "Shoulder surfing defence for recall-based graphical passwords", ACM Proceedings of the Seventh Symposium on Usable Privacy and Security. SOUPS'11, pp. 6:1-6:12, 2011

[139] H. Zhao; X. Li; "S3PAS: A Scalable Shoulder-Surfing Resistant Textual-Graphical Password Authentication Scheme," 21st International Conference on Advanced Information Networking and Applications Workshops, 2007, AINAW '07, vol.2, pp.467-472, 21-23 May 2007

Appendix A

The following table lists common authentication methods used in online environments.

Table 5: A list of authentication methods used in online environments. We provided the evaluations for each authentication type. The✓ means the authentication method satisfies the criteria/feature, ✓× means the authentication method partially satisfies the criteria/feature, and × means the authentication method does not satisfy the criteria/feature.

Authentication	Usability	Does not use extra devices	Performance / Accuracy	Addresses Key-logging attacks	Addresses Shoulder-surfing attacks	Addresses MITB attack	Overall Security
Text Password [123]	✓	✓	✓	×	×	×	✓×
Knowledge Based Questions [123]	✓	✓	✓	×	×	×	✓×
Graphical Password [6]	✓×	✓	✓×	✓	✓×	×	✓×
IP-Geolocation [30]	✓	✓	✓	✓	✓	×	✓×
Mobile-based SMS [30]	✓×	×	✓	✓	✓	×	✓×
OTP tokens [30]	✓×	×	✓	✓	✓	×	✓×
Grid based [31]	✓×	×	✓	✓×	✓×	×	✓×
Digital Certificates (Stored on PC, USB, or smartcard) [30]	✓×	×	✓	✓	✓	×	✓×
HAPTICS [70]	✓	×	✓×	✓	✓×	×	✓×
BIOMETRICS Physiological [55]	✓×	×	✓×	✓	✓	×	✓×
BIOMETRICS Behavioral [56]	✓	✓	✓×	✓	✓×	×	✓×

There is no authentication method that works for all types of applications. Each authentication method has its pros and cons. A combination of authentication methods could enhance security.

Appendix B

Proposed schemes

NOTE: A FORMAL SECURITY ANALYSIS AND A USABILITY CASE STUDY ARE REQUIRED TO DETERMINE WHETHER THE PROPOSED SCHEMES PROVIDE ENHANCED SECURITY AND/OR USABILITY TO CURRENT GRAPHICAL PASSWORD SCHEMES.

We propose 'Shift-based' and 'Grid Mapping' recognition-based graphical password schemes.

'Shift-based' scheme

The proposed 'Shift-based' scheme provides a new *approach* for the design of shoulder surfing resistant graphical password schemes. The proposed scheme differentiates itself from other graphical password schemes by being dynamic. It dynamically moves images for every user's action in order to resist shoulder surfing attacks. The scheme steps are game-oriented. The aim is to provide a game-oriented mechanism to login. In most shoulder surfing resistant schemes, the user clicks near the pass-images or in the middle of pass-images, however in the proposed scheme the user *does not* click anywhere near the displayed images or the pass-images. The user clicks take place from 3 buttons which are *outside* the displayed frame of images. Moreover, the user inputs are different for every authentication session. There are few shoulder surfing resistant graphical password schemes. The design of graphical password schemes that resist shoulder surfing attacks, including video recording and spyware (screen-capture), is a research area that still needs to be addressed. The concept from the proposed 'Shift-based' graphical password scheme can be developed into further new dynamic graphical password schemes that are game-oriented and resistant to shoulder

surfing attacks. The proposed *approach* can be used in future research to develop robust shoulder resistant graphical password schemes.

'Grid-mapping-based' scheme

The proposed 'Grid-mapping-based' scheme is a recognition-based graphical password scheme with challenge response protocol, which requires mental processing. The user has to recognize pass-images then do some steps. The scheme provides two-factor authentication and aims to resist shoulder surfing attacks. It combines knowledge-based, token-based grid-card, and process-based authentication. For knowledge-based, the user has to recognize pass-images. For token-based, the user has a dynamic paper-based *grid-card*. For the process-based, the user has to map numbers to images and images to numbers. The grid-card is paper-based card, unlike event-based and time-based OTP tokens that rely on electronic devices, where a device can break, malfunction, or the battery could die [29]. The proposed grid-card has two sides and is the size of a credit card when folding it. One side of the card is 'Numbers-Mapping-grid', and on the other side of the card, it is 'Images-Mapping-grid'. Since each grid box maps to a different grid box located at the other side of the card or at the displayed screen, the attacker has to guess the number or image of the mapping. The scheme aims to resist spyware. The spyware is located at the computer and the attacker does not have the grid-card. For each pass-image, there are 4 mapping steps. For each mapping step, there are 100 possibilities; making it 10^8 possibilities. Since the spyware only records what displayed on the screen and the actions of the user, it is very complex for the attacker to obtain the pass-images. By using the grid-card, each pass-image displayed on the screen is mapped four times to reach a random image on the screen where the user has to click on it. Since the attacker can only observe what was clicked on the screen, he/she still has to guess the 4 mappings (10^8 possibilities) to obtain the pass-image. The proposed scheme uses the concept of paper-based grid-card and the concept of grid mapping. The proposed *approach* can be used in future research to develop a robust graphical password scheme that addresses spyware.

Shift Based Scheme

The 'Shift-based' scheme provides a new approach to graphical passwords by dynamically moving objects for every user's action. The aim is to resist shoulder surfing attacks and makes the user feel like playing a *game* meanwhile shifting rows and columns.

Scheme Design

The proposed scheme is a recognition-based graphical password scheme. During registration, the user selects his/her N pass-images from a set of M images. During authentication, the scheme randomly displays the set of M images. The user has to find his/her pass-images and then shift rows and columns to match pass-images together. The authentication steps are performed in a game-oriented way. The values of N pass-images and the set of images M can vary depending on the application. In this work, we will have the values for N, and M to be 2, and 100, respectively. Figure-31 below displays M images in a RxC grid, where R is the number of rows, and C is the number of columns.

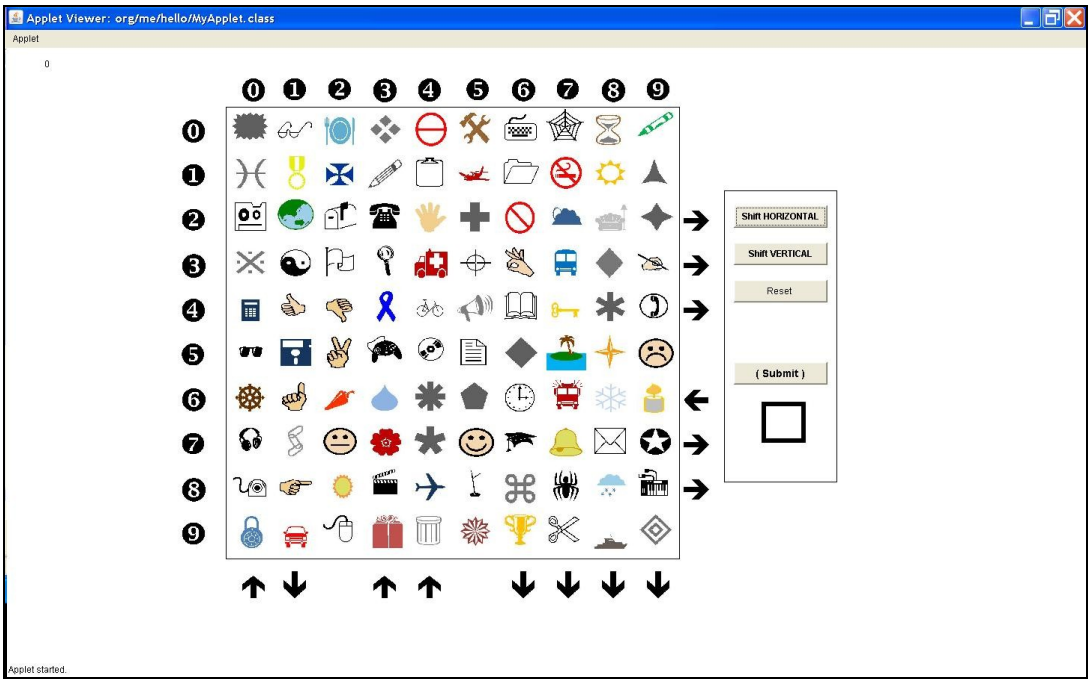


Figure 31: Scheme displaying M images, where M is 100 images.

From the displayed M images on the screen, the user has to perform the following steps:

- Find the two pass-images, 'pass-image-1' and 'pass-image-2'
- Mentally keep the locations of the pass-images
- Shift rows and columns to match the pass-images horizontally ('pass-image-1' to the left of 'pass-image-2') by clicking on 'Shift HORIZONTALLY', 'Shift VERTICALLY', and 'Reset' buttons in a minimum way in order to match the two pass-images together.
- Click Submit

The buttons to perform the match are:

- 'Shift HORIZONTALLY': Clicking on 'Shift HORIZONTALLY' button will shift rows horizontally. Each row will either shift one step to the right, one step to the left, or stay still. The arrow direction located at the right side of each row indicates the horizontal shift direction, see figure-32. A 'RIGHT-Arrow' (→) will shift the row one step to the right. A 'LEFT-Arrow' (←) will shift the row one step to the left. And 'No-Arrow' () will not shift the row in any directions.

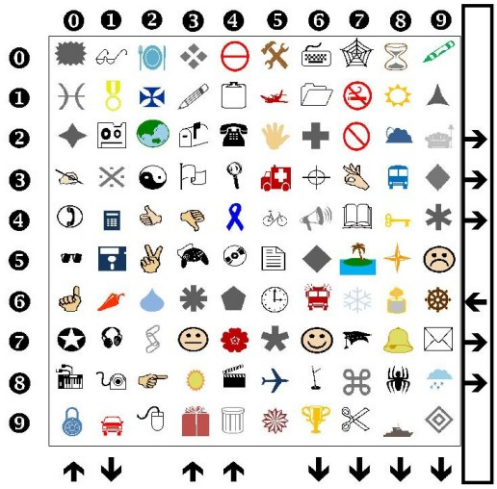


Figure 32: The black rectangle shows the horizontal arrows directions.

- 'Shift VERTICALLY': Clicking on 'Shift VERTICALLY' button will shift columns vertically. Each column will either shift one step upward, one step downward, or stay still. The arrow direction located below each column indicates the vertical shift direction, see figure-33. An 'UP-Arrow' (↑) will shift the column one step upward. A 'DOWN-Arrow' (↓) will shift the column one step downward. And 'No-Arrow' () will not shift the column in any directions.

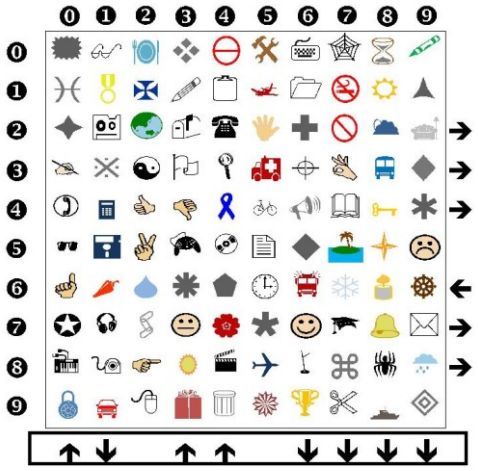


Figure 33: The black rectangle shows the vertical arrows directions

- 'Reset': Clicking 'Reset' button will change the directions of all Horizontal and Vertical arrows to random horizontal and vertical directions.

Scheme Scenario

This section will provide more details about the scheme by providing a simple scenario. In the following scenario, the number of pass-images N is 2, and the number of displayed images M is 100. The user's pass-images are:

'Pass-image-1': Earth icon, and 'Pass-image-2': Clock icon

During authentication, the user performs the steps explained in the previous section. From the displayed M images, the user finds his/her pass-images. Then in a game-oriented way, the user has to match them horizontally together.

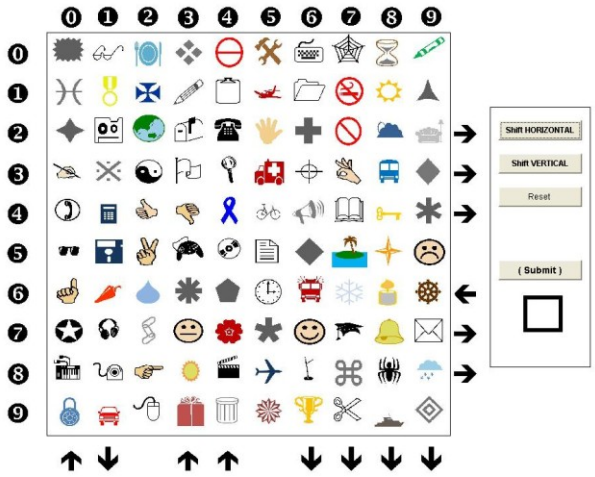


Figure 34: Displayed M images

From the displayed images, the user finds his/her pass-images. Figure-35 shows the pass-images in black boxes.

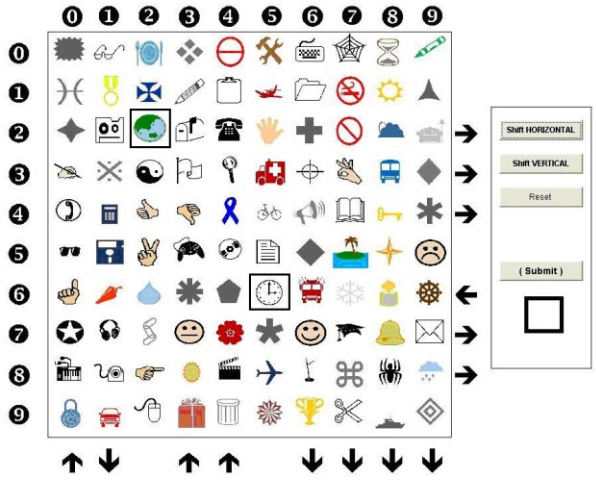


Figure 35: The user pass-images are shown in black boxes at locations 22, 65

The user has to match his/her pass-images (earth and clock icons) together. The clock icon (‘Pass-image-2’) has to be to the right side of the earth icon (‘pass-image-1’). The clock icon is located at (row ‘6’, column ‘5’), and the earth icon is located at (row ‘2’, column ‘2’), see figure-39. The arrow direction at row ‘2’ (earth icon row) is ‘RIGHT-Arrow’ and the arrow direction at row ‘6’ (clock icon row) is ‘LEFT-Arrow’. Clicking on ‘Shift HORIZONTALLY’ will shift row ‘2’ one step to the right, and will shift row ‘6’ one step to the left. The user does not have to click on ‘Reset’ button since the pass-images arrow directions will move the pass-images closer together. Therefore, the user clicks ‘Shift HORIZONTALLY’.

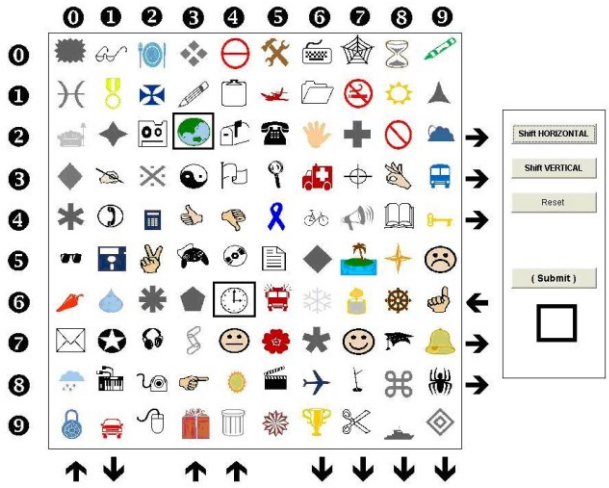


Figure 36: Displayed M images after the user clicks on 'Shift HORIZONTALLY'. The pass-images are shown in black boxes.

By clicking 'Shift HORIZONTALLY', the pass-images new locations are: 23, and 64, for earth, and clock icons, respectively. The clock icon is to the right side of the earth's icon. Note that all rows shifted according to the arrow directions located at the right side of each row. No arrow at the right side of the row resulted in the row to stay unmoved.

The user now has to shift columns in a way to match the pass-images. By clicking 'Shift VERTICALLY', the pass-images will both move upward since the vertical arrow direction below each column of the pass-images are pointing upward. In this way, the pass-image will not match or get close to each other, see figure-36. The user clicks on 'Reset' button to get new directions for the arrows.

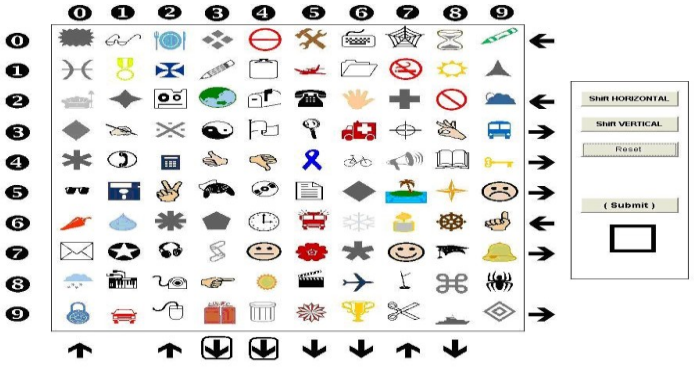


Figure 37: New arrow directions after the user clicks on 'Reset' button. Note that the new vertical directions for pass-images columns are both pointing downward.

As shown in Figure-37, the new vertical arrows directions for the pass-images are both pointing downward. The user has to click on ‘Reset’ button again to be able to shift columns ‘3’, and ‘4’ in a way to match the pass-images. The user clicks on ‘Reset’ button one more time.

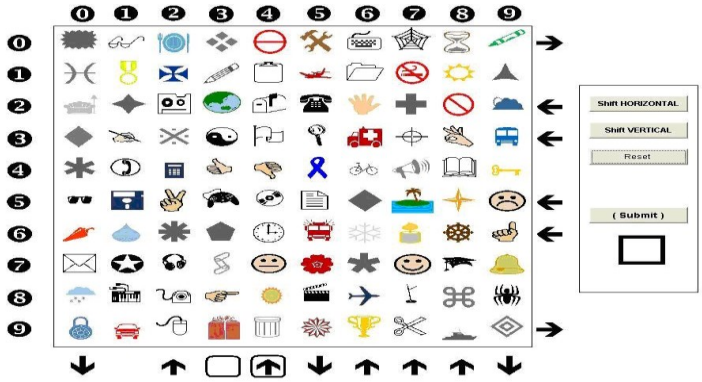


Figure 38: New arrow directions, after the user clicked on ‘Reset’ button. The new vertical directions for pass-images columns are shown in black boxes.

As shown in figure-38, the new vertical arrow direction for column ‘3’ is ‘No-arrow’, and the new vertical arrow direction for column ‘4’ is ‘UP-arrow’. By clicking on ‘Shift VERTICALLY’, the pass-images will get closer together vertically. The user clicks on ‘Shift VERTICALLY’ button.

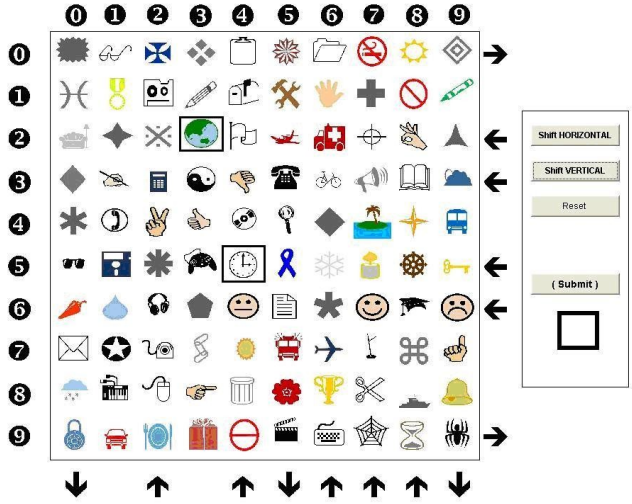


Figure 39: Displayed M images after the user clicks on ‘Shift VERTICALLY’. The pass-images are shown in black boxes.

To match the two pass-images, the user has to click on ‘Shift VERTICALLY’ three more times.

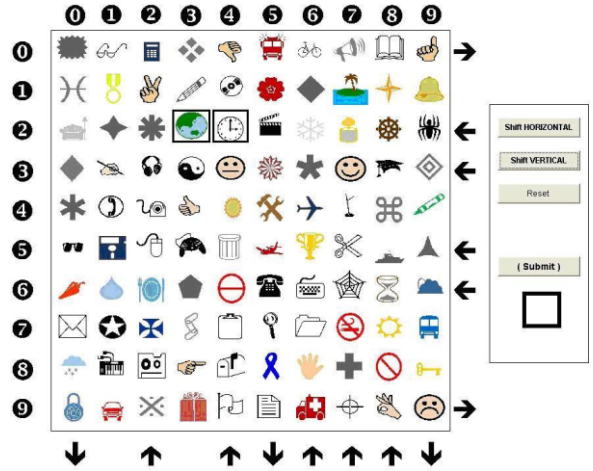


Figure 40: Displayed M images after the user clicks on ‘Shift VERTICALLY’ three more times. The pass-images are shown in black boxes.

As shown in figure-40 above, the two pass-images matched each other. The scheme tracks every movement and knows that there is a match. However, the user can perform couple more random shifts in order to avoid submitting images that have the two-pass-images next to each other. In this scenario, the user then randomly clicks on ‘Shift VERTICALLY’ two more times, then clicks ‘Reset’ button, then click on ‘Shift HORIZONTALLY’ one time. Finally the user clicks on ‘Submit’ button.

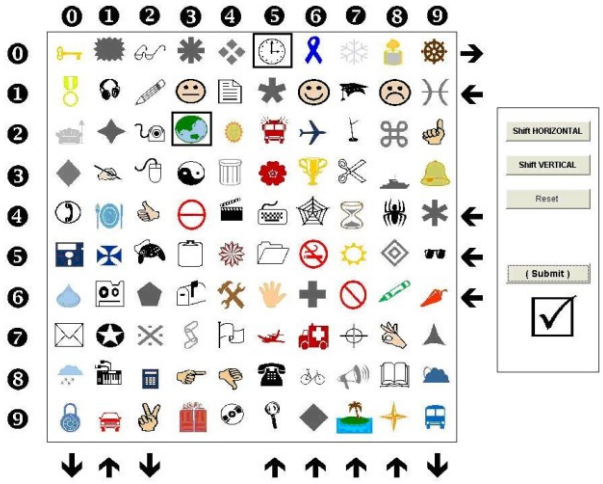


Figure 41: Final positions of M images. The two-pass images are shown in black boxes.

After the user clicked 'Submit' button, the user is authenticated since the two pass-images made a match during the authentication session.

Choosing a pair of images: The aim of the scheme is to resist shoulder surfing attacks. To resist against video recorded sessions and spyware (advanced screen-capture), we can match different pair of images for every authentication session. We are working on a technique called *'Get a pair of images from 4 pass-images'* technique that allows us to have a different pair of images to match for every authentication session. Instead of matching the pass-images, we can match two images that are chosen every time depending on the locations of the 4 pass-images. Once a user applies such a technique, the user can then apply the proposed 'Shift-Based' graphical password scheme. Applying the technique with 'Shift-Based' graphical password will make it difficult to get the password by shoulder surfing attacks, video recorded sessions, and spyware (advanced screen-capture).

Usability and Security Factors

Usability

For usability, factors to address are (i) easy to use, (ii) time to perform, (iii) and easy to remember. For ease of use, the scheme aims to be easy to use. For time to

perform, the scheme takes up to few minutes to login, however in a game-oriented way. And for ease to remember, the scheme aims to be easy to remember since the user has to recognize a few pass-images from a set of M images. A user case study on 20-30 users will take place to evaluate the usability of the scheme on computers, tablets, and mobile phones.

Password Space

The password space for this scheme is C(N,K).

$$\binom{N}{K}$$

Choosing K pass-images from N images.

For N = 100, and K = 2, the password space is approximately 5×10^3

The password space is very small from this scheme. The proposed scheme can be an add-on to a text based password to add additional layer and to have a larger password space.

Observation Attack

There are many capture based attacks that graphical passwords have to address: (i) shoulder surfing, (ii) eavesdropping, and (iii) spyware [123].

Shoulder Surfing

Shoulder surfing attack takes place in various ways [37]; (i) mainly by an observer/attacker observing the user meanwhile the user is entering their password; and (ii) user is being recorded by an electronic device such as video cameras, and phones.

For (i), the scheme aims to resist against an observer. For (ii & iii), the scheme aims to be resistant to video recording and spyware for a couple of successful login sessions, without giving hints about the password to the observer/attacker. Analyzing the technique and other possible techniques, it was found that such techniques are not very strong against many recorded successful logins.

Eavesdropping

To assure prevention of eavesdropping, HTTPS should be used to provide security and confidentiality between the client and the server [123].

Spyware

The scheme does not address spyware.

Scheme Overview

The scheme aims to resist shoulder surfing attacks by having a dynamic movement of images for every action clicked by the user. The scheme can be an add-on to text-password to have a strong password size. The scheme benefits come with few drawbacks. To get authenticated, it takes from less than a minute to few minutes. However, the scheme steps are game-oriented. The user has to match two images together by shifting rows and columns. The scheme aims to resist key-logging and shoulder surfing attacks. In most shoulder surfing resistant schemes, the user clicks near the pass-images or in the middle of pass-images, however in the proposed scheme the user does not click anywhere near the displayed images or the pass-images. The user clicks take place from 3 buttons which are outside the displayed frame of images. Moreover, the user inputs are different for every authentication session.

For the proposed graphical password, mental processing (something the user processes) is required. The user has to recognize pass-images and then perform some steps. The scheme resists simple shoulder surfing, but is not strong against spyware and video recording for so many recorded successful logins. Moreover, the scheme benefits come with drawbacks. It takes a few minutes to get authenticated, which is longer than other graphical passwords schemes. Moreover, the user has to learn steps in order to enter the challenge response. However, the scheme is shift-based to match pass- images together. The steps are game-oriented, so the user feels like playing a game meanwhile

getting authenticated. The proposed 'Shift-Based' scheme is strong against key-logging, and aims to resist simple shoulder surfing. Nonetheless, the login inputs are different on every login session.

Scheme Implementation

The proposed 'Shift-based' graphical password scheme is implemented as Java Applets using NetBeans IDE 7.0. None of the images and icons belongs to us. The images and icons are taken from Microsoft Word several fonts including Webdings, Wingdings, Wingdings 2, and Wingdings 3. Paint and colors were added to some images.

Future work

We will create the *'Get a pair of images from 4 pass-images'* technique in order to match different pair of images for every login. Once completed, security analysis and usability surveys will be conducted. Furthermore, the concept from the 'Shift-based' scheme can be developed into further new dynamic graphical password schemes in order to resist shoulder surfing attacks.

Conclusion

There are few shoulder surfing resistant graphical password schemes. Further research is required in shoulder surfing resistance schemes. We propose a new approach for the design of shoulder surfing resistant graphical password schemes. The 'Shift-based' scheme dynamically moves images for every user's action in order to resist

shoulder surfing attacks. The scheme steps are game-oriented. The aim is to provide a game-oriented mechanism to login. Furthermore, the 'Shift-based' scheme provides a new approach to graphical password schemes in order to resist shoulder surfing attacks. The design of graphical password schemes that resist shoulder surfing attacks, including video recording and spyware (screen-capture), is a research area that still needs to be addressed.

Grid Mapping Based Scheme

In this section, we propose a 'Grid-Mapping-based' scheme that is based on two factor authentication.

Scheme Design

The proposed 'scheme is a recognition-based graphical password with challenge response protocol, which requires mental processing. The user has to recognize pass-images then do some steps. The scheme is a multi-factor authentication. It combines knowledge-based, token-based, and process-based authentication. For knowledge-based, the user has to recognize pass-images. For token-based, the user has a dynamic *grid-card*. For the process-based, the user has to map numbers to images and images to numbers. The grid-card can be e-grid-card or physical grid-card. In this work, we will use the physical grid-card. Unlike event-based and time-based OTP tokens that rely on electronic devices, where a device can break, malfunction, or the battery die [29]. The proposed grid-card is a two sideway paper-based card, which has the size of a credit card when folding it. One side of the card is 'Numbers-Mapping-grid' that has a CxR grid, see figure-42. Where C is the number of columns, and R is the number of rows. Starting from the top-left, each cell has a unique number in the range of [0, M-1].

Furthermore, each user will have a random set of numbers in their Numbers-Mapping-grid' card.

X	0	1	2	3	4	5	6	7	8	9
0	90	36	81	32	45	68	20	86	75	47
1	7	58	17	26	39	70	13	82	0	53
2	65	11	31	44	61	2	77	25	43	95
3	23	56	49	83	52	40	96	64	85	21
4	42	80	1	88	9	76	19	27	14	5
5	37	66	71	94	48	91	62	97	28	54
6	78	30	50	35	60	3	57	10	18	67
7	29	8	73	55	98	79	92	24	34	15
8	93	4	12	33	16	22	46	41	99	69
9	51	38	87	74	59	72	6	84	63	89

Figure 42: An example of a 'Numbers-Mapping-grid' card, for M equals 100

The other side of the card is 'Images-Mapping-grid' that has a CxR grid, see figure-43. Where C is the number of columns, and R is the number of rows. Starting from the top-left, each cell has an image and a number starting from (0) to (M-1). Furthermore, each user's grid card will have the images ordered randomly in their 'Images-Mapping-grid' card.

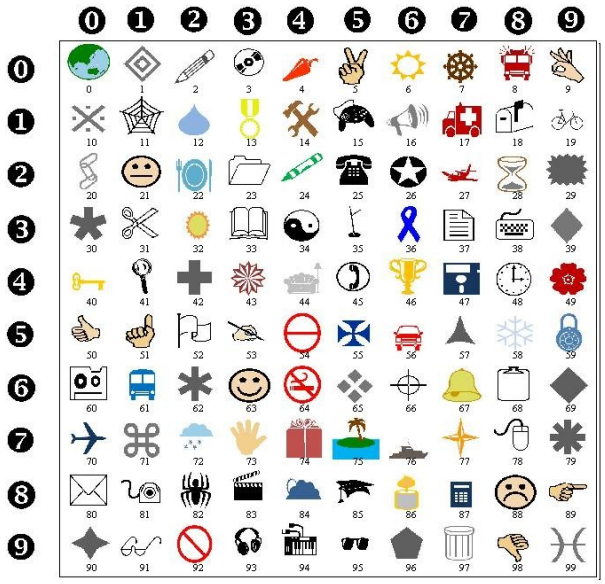


Figure 43: An example of 'Images-Mapping-grid' card, for M equals 100

During registration, the user selects N images as pass-images from a set of M images. The values of N and M can vary. We suggest N to be 3 pass-images, and M to be 100 images. During authentication, the scheme randomly displays the set of M images in a CxR grid, see figure-44.

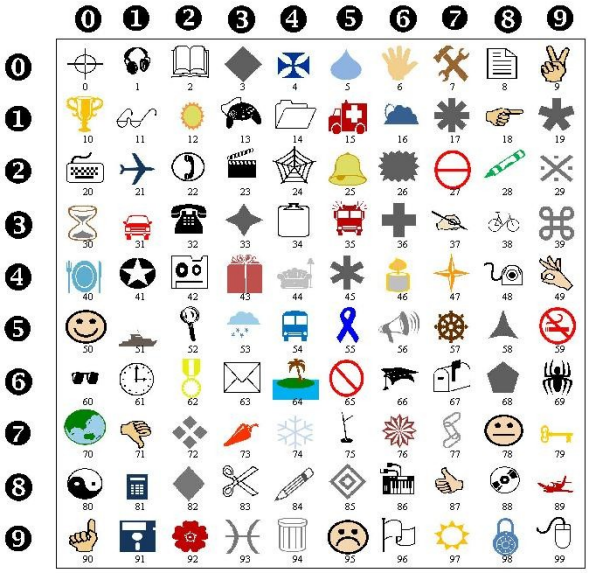


Figure 44: A set of M randomly displayed images, where M equals 100.

Where C is the number of columns, and R is the number of rows. Starting from the top-left image, each image has a location number. The images locations numbers start from 0 to 99, for M equals 100 images. During login session, the user has to perform the following steps:

Steps:

- The user has to find the N pass-images from the set of M images.
- From top-left, the user looks for N pass-images
- For each pass-image found:
    - The user finds the location ‘c’ of the pass-image, where ‘c’ is between 0 and 99:

- User applies **'NumbersGrid_MappingTo_ImagesGrid'** (see figure-45):
    - The user looks in the number-based grid-card to find location 'c'
    - The user finds number 'd' inside the cell content of location 'c'
    - The user checks the displayed frame to find the image at location 'd'
    - The user finds image 'i1'
    - The user looks in the 'image-based grid-card' to find the location of image 'i1'
    - In 'image-based grid-card', location of image 'i1' is 'e'
    - On the displayed frame, the user *clicks* on the image at location 'e'.
- User applies **ImagesGrid_MappingTo_NumbersGrid** (see figure-46)
    - The user looks in the 'image-based grid-card' to find location 'c'
    - The user finds image 'i2' inside the cell content of location 'c'
    - User checks the displayed frame to find location of image 'i2'
    - User finds location 'f'
    - User looks in the 'number-based grid-card' to find cell content at location 'f'
    - In number-based grid-card, the user finds number 'g' at location 'f'
    - On the displayed frame, the user *clicks* on image at location 'g'
- User repeat steps for the other pass-images

- The user clicks on Submit button

NumbersGrid_ MappingTo_ImagesGrid (summary)**:**

Given: Pass-Image @ location → ‘c’,

The following are the steps, as shown in figure-45:

Step (1): NumbersGrid @ location ‘c’ → number ‘d’

Step (2): Display Frame @ location ‘d’ → image ‘i1’

Step (3): ImagesGrid @ image ‘i1’ → number ‘e’

Step (4): Display Frame @ location ‘e’ → CLICK on the image at location ‘e’

Therefore:

[NumbersGrid (‘c’) → ‘d’ → Find it in Displayed Frame → @(‘i1’) → Get from ImagesGrid (‘i1’) → ‘e’]

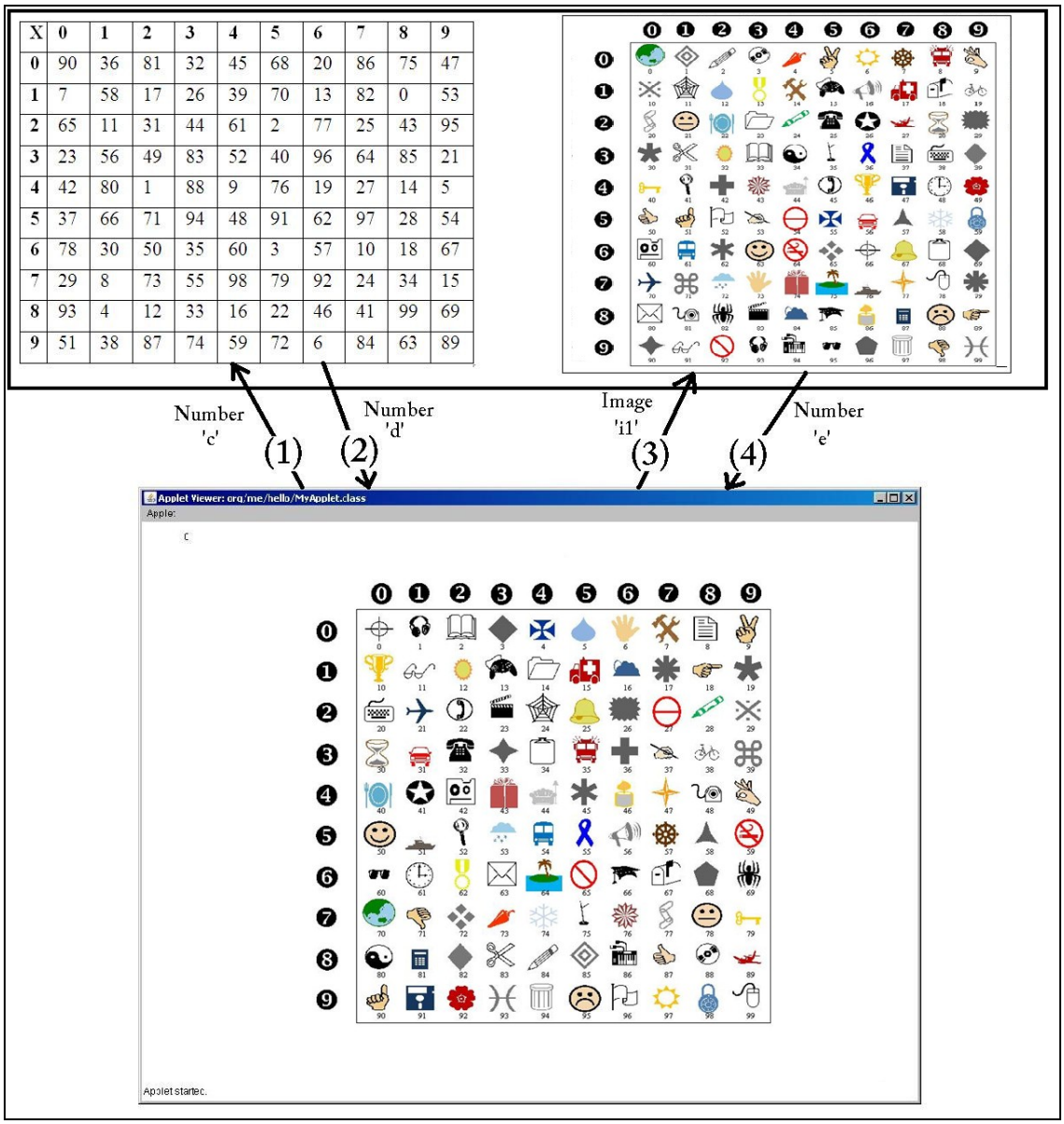


Figure 45: Steps of NumbersGrid_ MappingTo_ImagesGrid

ImagesGrid_ MappingTo_NumbersGrid (summary)**:**

Given: Pass-Image @ location → 'c',

The following are the steps, as shown in figure-46:

Step (1): ImagesGrid @ location 'c' → image 'i2'

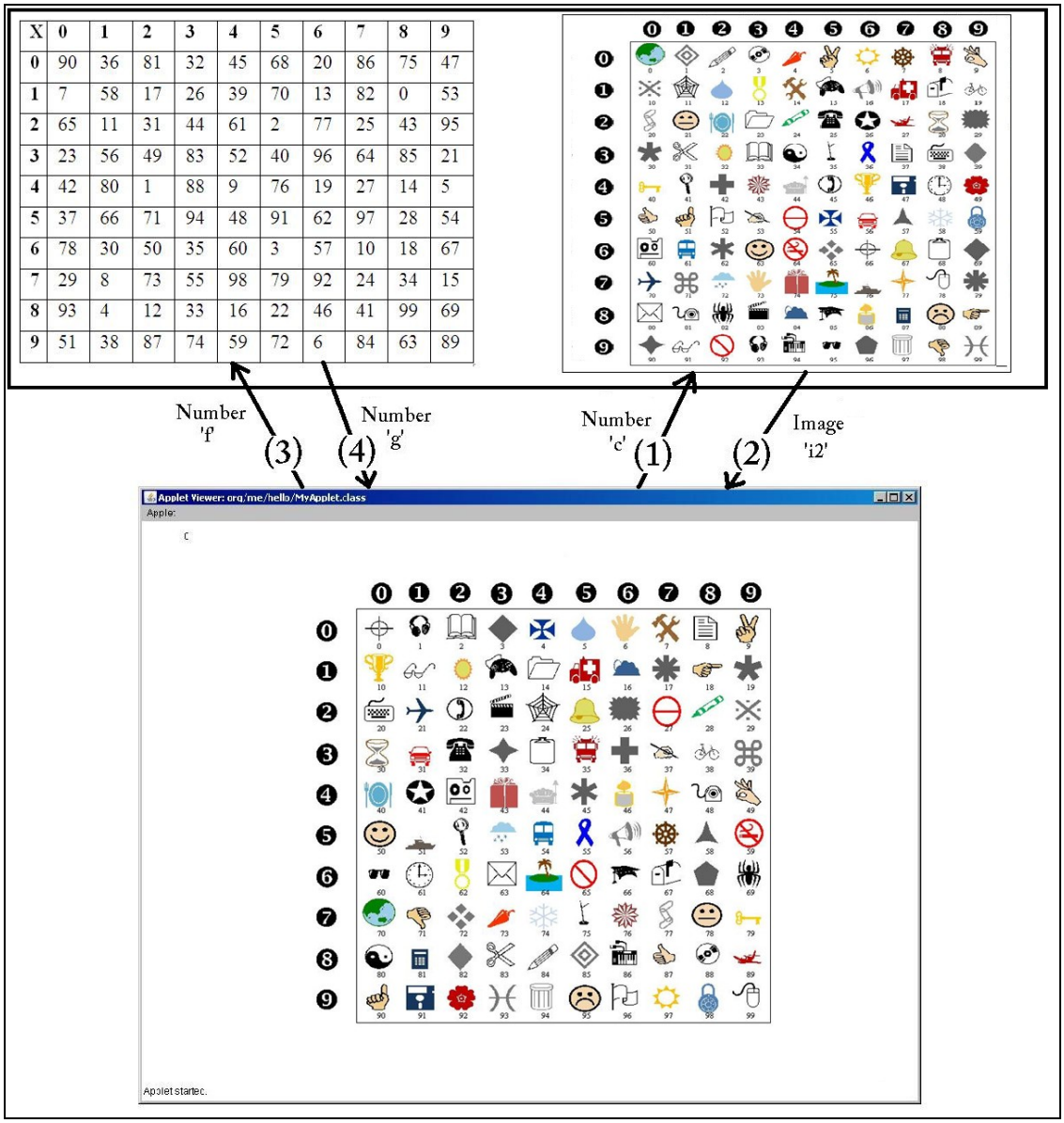
Step (2): Display Frame @ image ‘i2’ → number ‘f’

Step (3): NumbersGrid @ location ‘f’ → number ‘g’

Step (4): Display Frame @ location ‘g’ → CLICK on the image at location ‘g’

Therefore,

[ImagesGrid (‘c’) → ‘i2’ → Find it in Frame → @(‘f’) → Get from NumbersGrid (‘f’) → ‘g’]



X	0	1	2	3	4	5	6	7	8	9
0	90	36	81	32	45	68	20	86	75	47
1	7	58	17	26	39	70	13	82	0	53
2	65	11	31	44	61	2	77	25	43	95
3	23	56	49	83	52	40	96	64	85	21
4	42	80	1	88	9	76	19	27	14	5
5	37	66	71	94	48	91	62	97	28	54
6	78	30	50	35	60	3	57	10	18	67
7	29	8	73	55	98	79	92	24	34	15
8	93	4	12	33	16	22	46	41	99	69
9	51	38	87	74	59	72	6	84	63	89

Figure 46: Steps of ImagesGrid_ MappingTo_NumbersGrid

Therefore, for each pass-image found, the user has to map the following in order to get the values of ‘e’ and ‘g’:

NumbersGrid (‘c’) → ‘d’ → Find it in Frame → @(‘i1’) → Get from ImagesGrid (‘i1’) → ‘e’

ImagesGrid (‘c’) → ‘i2’ → Find it in Frame → @(‘f’) → Get from NumbersGrid (‘f’) → ‘g’

After simple training, the user can perform the steps.

Scheme scenario

The user has a grid-card that have a table on each side. One side has a ‘number-based grid-card’ and the other side has ‘image-based grid-card’, as shown in figure-47.

X	0	1	2	3	4	5	6	7	8	9
0	90	36	81	32	45	68	20	86	75	47
1	7	58	17	26	39	70	13	82	0	53
2	65	11	31	44	61	2	77	25	43	95
3	23	56	49	83	52	40	96	64	85	21
4	42	80	1	88	9	76	19	27	14	5
5	37	66	71	94	48	91	62	97	28	54
6	78	30	50	35	60	3	57	10	18	67
7	29	8	73	55	98	79	92	24	34	15
8	93	4	12	33	16	22	46	41	99	69
9	51	38	87	74	59	72	6	84	63	89

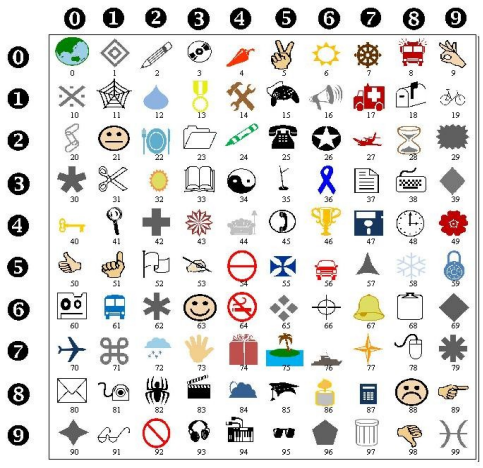


Figure 47: The user grid-card having ‘number-based grid-card’ on one side and ‘image-based grid-card’ on the other side.

The user pre-selected pass-images are: coffee cup, pencil, and apple.

Clock:  **, Bell:**  **, Airplane:** 

During authentication, the scheme provides the user with a set of M images randomly ordered each time. The user has to find the N pass-images.

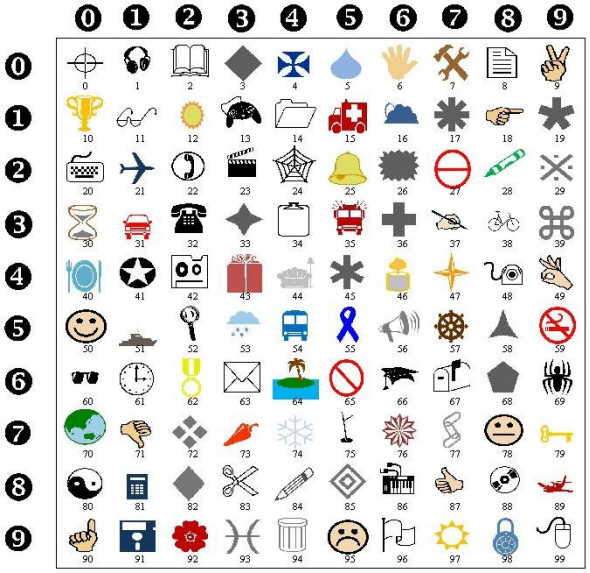


Figure 48: A set of 100 randomly displayed images

The user pre-selected pass-images: clock, bell, and airplane are located at 61, 25, and 21 respectively. See figure-48.

For Pass-Image (clock):

It is located at number (61):

NumbersGrid (‘61’) → ‘30’ → Find it in Frame → @(‘Sand Timer’) → Get from ImagesGrid (‘Sand Timer) → **‘28’**

ImagesGrid (‘61’) → ‘Bus’ → Find it in Frame → @(‘54’) → Get from NumbersGrid (‘54’) → **‘48’**

For Pass-Image (bell):

It is located at number (25):

NumbersGrid (‘25’) → ‘2’ → Find it in Frame → @(‘Book’) → Get from ImagesGrid (‘Book’) → **‘33’**

ImagesGrid (‘25’) → ‘Telephone’ → Find it in Frame → @(‘32’) → Get from NumbersGrid (‘32’) → **‘49’**

For Pass-Image (airplane):

It is located at number (21):

NumbersGrid (‘21’) → ‘11’ → Find it in Frame → @(‘Eyeglasses’) → Get from ImagesGrid (‘Eyeglasses’) → **‘91’**

ImagesGrid (‘21’) → ‘Smile Face’ → Find it in Frame → @(‘50’) → Get from NumbersGrid (‘50’) → **‘37’**

Therefore, the user has to click on 6 images at locations: **28, 48, 33, 49, 91,** and **37**.

Usability and Security Factors

Usability

For usability, factors to address are (i) easy to use, (ii) time to perform, (iii) and easy to remember. For ease of use, the scheme is not easy to use at the beginning. It requires the user to memorize couple of instructional steps. For time to perform, the scheme can take up to few minutes. Furthermore, for ease to remember, the scheme aims to be easy to remember since the user has to recognize a few pass-images from a set of M images. A user case study on 20-30 users will take place to evaluate the usability of the scheme on computers, tablets, and mobile phones.

Password Space

The password space for this scheme is N^K.

N is the number of images, and K is the number of mapped-imaged to be clicked.

For N = 100, and K = 6, the password space is approximately $100^6 = 10^{12}$

The password space is close to the text-based password space which is approximately 10^{14}.

Observation Attack

There are many capture-based attacks that graphical passwords have to address: (i) shoulder surfing, (ii) eavesdropping, and (iii) spyware [123].

Shoulder Surfing

Shoulder surfing attack takes place in various ways [37]; (i) mainly by an observer/attacker observing the user meanwhile the user is entering their password; (ii) user is being recorded by an electronic device such as video cameras, and phones. For (i), the scheme aims to resist against an observer. And for (ii), the scheme aims to be resistant to video recording. The scheme requires the user to have the NumberGrid and ImageGrid Card. Since each grid location maps to a different grid location, the attacker has to guess the number or image of the mapping. For each mapping step, there are 100 possibilities. For each pass-image, the end of the process is two clicks on 2 of the 100 images.

Eavesdropping

To assure prevention of eavesdropping, HTTPS must be used to provide security and confidentiality between the client and the server [123].

Spyware

The scheme aims to address spyware. A full security analysis will be provided to verify the claim.

Scheme Overview

The ‘Grid-Map-based’ scheme is based on two factor authentication. Its benefits come with drawbacks. It requires the user to remember instructional steps, and can take long time to perform. However, it provides two-factor authentication and aims to resist shoulder surfing attacks.

Appendix C

The following section provides a method to detect and recognize a human ear from a video. We can add ear detection and recognition to have a multimodal biometric system based on the fusion of face, voice, and ear biometrics.

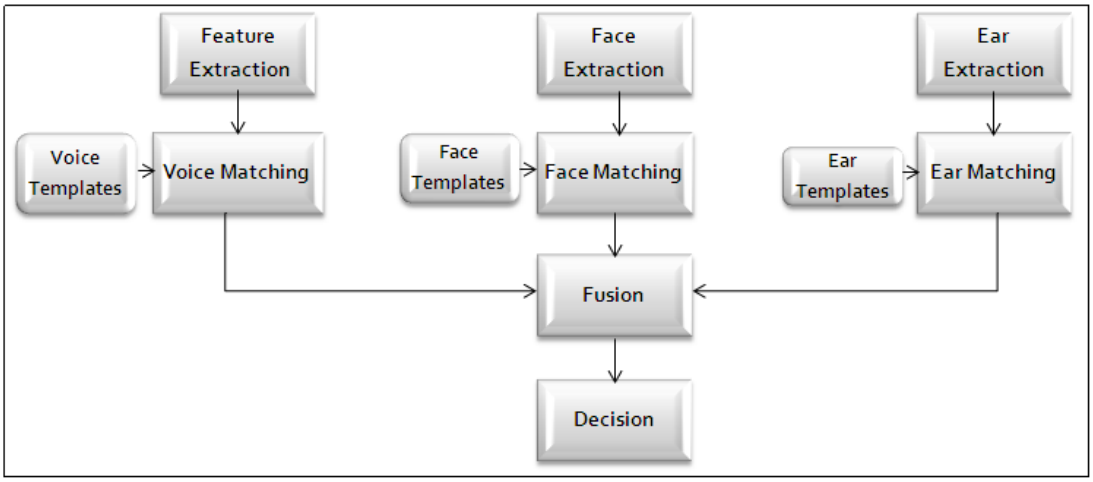


Figure 49: System Modules

Ear Detection and Recognition

An ear detection system can detect the user's right or left ear [10]. When the user looks aside at angle 20 degrees or more from a straight pose, the system shall detect the ear. OpenCV can be used for ear detection by using [100]. [100] is a technique that is based on the Viola-Jones method to detect ears in real-time. The ear detector can detect ears in general, right ears, and left ears. For the right ear detection, '12x12 Right ear detector' provided by [100] can be used. For the left ear detection, '12x12 Left ear detector' provided by [100] can be used.

The detected ear at a specific face pose is compared to a template ear that has approximately the same face pose. This way, the detected ear and the template ear have approximately the same pose. Therefore, face pose estimation is required. We used a face pose approximation method that is based on the face and eyes detection. During the training phase, the user's ear samples can be extracted from different face poses. The face poses can be between –70 degrees to +70 degrees from a frontal face pose. The method approximates the face pose based on the right and left eyes detection boxes around each eye. When the user has a frontal face pose, the sizes of the boxes around the right and left eyes are almost the same. If the user poses to their right side, the left eye box gets slightly bigger and the right eye box gets smaller until the user poses to a degree where the right eye box disappears. Vise-versa, if the user poses to the left side. The ratio between the sizes of the boxes can estimate the face pose. Therefore, the OpenCV [35] cascades were not only used to detect the users' right and left eyes, but also to provide face pose approximation. The latter can be used to provide the system with the user's face pose approximation in order to have the rotation/angle of the face pose of the test ear. Then, the detected ear is compared to an ear template that has approximately the same face pose. The cascade used is for right eye is the 'Tree-based 20x20 right eye detector' provided by [35]. Moreover, the cascade used is for left eye is the 'Tree-based 20x20 left eye detector' provided by [35].

For ear recognition, Eigenimage [75] approach can be used. The technique is well defined in [75, 120, 121]. If ear recognition is part of a multimodal biometric system, the computation results are sent to fusion module. A well-known one is Score level fusion [54]. We can then use the Min-Max Normalization [17, 54] and simple weighted sum rule [17, 21, 97, 129] as was done in chapter 3:

$$S_e = Score_e - min_{eScore} / max_{eScore} - min_{eScore}$$

Where, $Score_e$ is ear biometric score, max_{eScore} is maximum score for ear, min_{eScore} is minimum score for ear, and S_e is the normalized ear score.

The normalized score can have a similarity score or a dissimilarity score [17, 54]. To have a common domain, the biometric trait scores based on dissimilarity scores

are converted to similarity scores [17, 54]. To convert dissimilarity score to similarity score, the dissimilarity score is subtracting from 1 [54]. The conversion of each score S results in NS.

Applying the weighted sum rule results in [17, 21, 97, 129]:

$$NScore_{Total} = W_v * NS_v + W_f * N_f + W_e * NS_e$$

Where, $NScore_{Total}$ is total score from the biometric traits, W_v is weight for voice biometric, W_f is weight for face biometric, W_e is weight for ear biometric, NS_v is the normalized score for voice biometric, NS_f is the normalized score for face biometric, NS_e is the normalized score for ear biometric. Each of weights W is in range of [0, 1] and the total weights sum equals to 1 [17, 54].

$$W_v + W_f + W_e = 1$$

The fusion module will send the result of $NScore_{Total}$ to the decision module. At the decision module, the $NScore_{Total}$ can be compared to a threshold value to determine if the claimed identity is a genuine or an impostor. The threshold value has to be selected in order to increase the True-Positive (GAR) and decrease False-Positive (FAR).

For face detection, ear detection, eye detection, face pose approximation, ear recognition, and cascade detectors, OpenCV and Matlab open-source code can be used. For ear recognition, open-source EigenImage implementation can be used. Adding those components will enable us to detect and recognize a human's voice, face, and ear biometrics from a video.