Identity Theft, Computers and Behavioral Biometrics

Robert Moskovitch, Clint Feher, Arik Messerman, Niklas Kirschnick, Tarik Mustafić, Ahmet Camtepe, Bernhard Löhlein, Ulrich Heister, Sebastian Möller, Lior Rokach, Yuval Elovici

Abstract— **The increase of online services, such as eBanks, WebMails, in which users are verified by a username and password, is increasingly exploited by Identity Theft procedures. Identity Theft is a fraud, in which someone pretends to be someone else is order to steal money or get other benefits. To overcome the problem of Identity Theft an additional security layer is required. Within the last decades the option of verifying users based on their keystroke dynamics was proposed during login verification. Thus, the imposter has to be able to type in a similar way to the real user in addition to having the username and password. However, verifying users upon login is not enough, since a logged station/mobile is vulnerable for imposters when the user leaves her machine. Thus, verifying users continuously based on their activities is required. Within the last decade there is a growing interest and use of biometrics tools, however, these are often costly and require additional hardware. Behavioral biometrics, in which users are verified, based on their keyboard and mouse activities, present potentially a good solution. In this paper we discuss the problem of Identity Theft and propose behavioral biometrics as a solution. We survey existing studies and list the challenges and propose solutions.**

I. Introduction

THE increasing use of the internet presents many opportunities for exploitation through identity theft. Identity theft is frequently used by intruders to access web accounts through the internet. Currently users identify themselves often by a user-name and a password. The common use of the same password for many services increases the vulnerability whenever such information is being theft. Thus, an additional security means is required for confirming the identity. Potentially useful technologies for solving this can be found in biometrics.

Since the terrorist attack on September 11, 2001 there was a growing interest in the use of biometrics for identity verification. However, relying on such technology in the internet is very limited because of the requirement of dedicated hardware devices which are costly and often not available. Recently laptops come with a fingerprint verification device; however, this is still not popular enough and can not be used for user verification in web applications. Thus, a good way to verify users' biometric properties can rely on the interaction of the user with certain devices such as keyboard and pointing devices. User can be even verified based on the way he is using certain applications. Within the past three decades several studies were made in the use of keystroke dynamics for verification of users upon login and for free texts [1]. Recently the use of mouse for biometric verification was proposed [2]. The main advantage of this option is its availability with no additional cost; however, there are still several challenges which should be overcome in order to make it an operative technology.

Identity theft is a fraud, in which someone pretends to be someone else in order to steal money or get other benefits: from the more traditional financial crimes that ranges from loan, mortgage, credit card, commodities and services frauds, to money laundering, trafficking human beings, stock market manipulation and even breaches of national security or terrorism. According to the non-profit Identity Theft Resource Center (ITRC[1]), identity theft from a consumer perspective is sub-divided into four categories:

- *Financial identity theft* (using another's identity to obtain goods and services), for example a *bank fraud*.
- *Criminal identity theft* (posing as another when apprehended for a crime)
- *Identity cloning* (using another's information to assume his or her identity in daily life)
- *Business/commercial identity theft* (using another's business name to obtain credit)

In this paper we present the problem of identity theft in personal devices, such as personal computers and mobile devices, through which users frequently access their data on websites. Identity theft can be used to access local valuable information stored on personal computers or mobile devices, which will become more and more important with the increase in storage size and functionality. Another option is to access services provided through a network of computers, such as the internet, and intranet for organizations.

We refer to the option of biometrics, and behavioral biometrics in particular, through keystroke and mouse

Robert Moskovitch, Clint Feher, Lior Rokach and Yuval Elovici are with the Deutsche Telekom Laboratories at Ben Gurion Univeristy, Ben Gurion University, Be'er Sheva, 84105, Israel (e-mail: robertmo, clint, lior, elovici@bgu.ac.il).

Arik Messerman, Tarik Mustafić and Ahmet Camtepe are with the DAI-Labor at the Technical University of Berlin, DAI-Labor, , Ernst-Reuter-Platz 7, 10587 Berlin, Germany, (e-mail: arik.messerman, tarik.mustafić, ahmet.camtepe@dai-labor.de).

Niklas Kirschnick and Sebastian.Möller are with the Quality and Usability Lab, Deutsche Telekom Laboratories, TU Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany, (e-mail: niklas.kirschnik, sebastian.moeller@telekom.de).

Ulrich Heister and Bernhard Löhlein are with Deutsche Telekom, Laboratories, Innovation Development at Deutsche-Telekom-Allee 7, 64295 Darmstadt, Germany (e-mail: {ulrich.heister, bernhard.loehlein}@telekom.de

[1] http://www.idtheftcenter.org

dynamics to add an additional security layer for the devices and websites. We survey previous scientific studies, list the potential uses of this technology, and present a general framework. Finally we refer to the various challenges and open problems of this technology.

II. Computer Security through Biometrics

Within the last decades the use of electronic biometrics devices is increasing to better secure computer networks and services. Two types of biometrics are traditionally distinguished: *physiological* and *behavioral*. Physiological biometrics refers to physical measurements of the human body, such as *fingerprint*, *face*, *hand* (*palm*) *geometry* and *iris*. Physiological biometrics often relies on a snapshot (single moment) in which measurements of the users are scanned, however, this often relies on the assumption that physiological properties do not change very rapidly, so they can easily be exploited for identity theft.

Behavioral biometrics [3] relates to the specific behavior of a human (user) along time in performing some task, such as *signature writing*, *voice*, *keystroke dynamics* and others. The major difference among the physiological and the behavioral biometrics is the temporal aspect, which makes the latter harder to detect and imitate. As a consequence, behavioral biometrics was largely ignored for user verification in the past. Among the varying types of behavioral biometrics we are focusing in this paper on the use of keystroke and mouse dynamics in the task of user verification.

For the evaluation of biometric systems there are common measures, which we will be referring to throughout the paper: False Acceptance Rate (FAR), which is the rate an imposter could be verified or identified by the biometric method, and False Rejection Rate (FRR), which is the rate a legitimate user is rejected by the biometric system.

III. Behavioral Biometrics in Computers

Behavioral biometrics has the potential to verify users based on their interaction with the computer, through the keyboard and mouse, in several applications:

A. *Common Scenarios of Behavioral Biometrics*

There are several scenarios in which Behavioral Biometrics can be used in Computers:

- Log-in Verification

  Whenever the user logs to his local station, or to a service on the intranet or internet by typing a user-name and password - these are monitored and verified.
- Continuous Verification

  After the user performs login to the computer or to the web service, his entire interaction, through keyboard, mouse (and sometimes applications) activities are continuously monitored to verify that it remains him.
- Password Reset

  When the user has forgotten his credentials for login, the user is asked to perform a behavioral biometric verification process/task instead of contacting a telephone hotline or visiting the office of the administrator.

B. *Keystroke Dynamics*

Keystroke dynamics can be captured via several different features extracted from the typing rhythm of the user including: *latency* between consecutive keystrokes, *flight time*, *dwell time*, based on the key down/press/up events (as shown in Figure 1), overall typing speed, frequency of errors (use of backspace) and control keys (use of left/right shift). Systems do not necessarily employ all of these features; most of the applications measure often only *latencies* and *dwell time*. Features of keystroke sequences, often used for long texts verification, are typically extracted based on *di-graph*, *tri-graph* or (more generally) *n-graph* segments of the entire text. In these, the *latencies*, *intervals* and *flight time* are measured for each sequence of keystrokes. Verification methods are dedicated to verify users based on fixed (static) or variable (free) text inputs. Latter methods can be used in order to verify users continuously.

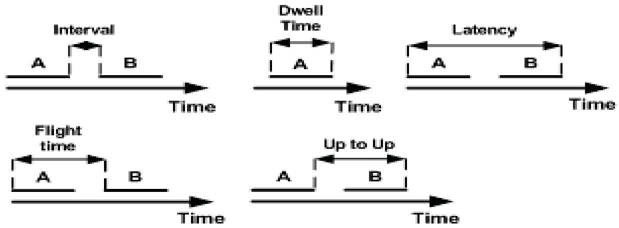

Fig 1: Keystroke metrics: *latency*, *interval*, *dwell time* and *flight time*.

Generally, typing pattern solutions can be divided in those, which analyze the user behavior during an initial login attempt and those, which determine the authentication state of a user permanently. Cho et al. [4] measure the delay between key presses and the *dwell* time that are then processed in a multilayer perceptron neural network in order to discriminate between the user and an imposter. Adjusting the threshold they achieved a FAR of 0.0% and FRR of approximately 1%. The same characteristics were already used in [5]. Lin et al presented similar results (FAR 0,0%, FRR 1.1%) based on the input of passwords with the length of six to eight characters.

Lau et al. [6] state that formerly used metrics for verification do not perform very well. Therefore, they propose four new metrics: *key press duration*, *relative key event order*, *relative keystroke speed* and classes of *shift key usage*. They evaluated every metric on its own and used a statistical analysis model for that. Revett et al. [7] analyzes keystrokes of the passphrase with a constant length of 14 characters for every user. They calculated a similarity measure to create a decision table and used this table to determine rules based on rough sets. With this method a 97% accuracy for a newly entered sample was achieved.

Bartmann et al. [8] patented a verification approach based on typing behavior. In this patent it is described which features can be used to verify users. Besides some of the ones mentioned earlier, they also propose to use the *overlapping of key presses*, i.e. one key is still held while the next one is pressed, which is actually a negative interval (fig1). Furthermore, the usage behavior of the shift keys is used. Another important aspect which is considered by Bartmann et al. is the fluctuation of the typing behavior and how to cope with these changes. Bartmann et al established a company, called Psylock[2], which verifies the user based on an identical typed string of around 49 characters, which is used for all the users. Psylock is used at the University of Regensburg to enable users reset their (lost) passwords [9] without the requirement of an administrator. Obviously, such a long password is not usable for daily use in websites and other services, in which it is often a unique 8 characters password.

Bergadano et al. [10] proposed to use relative duration times of n-graphs instead of absolute ones. That means, the graphs were sorted by their duration and then the distance between the single graphs is calculated and compared to other users typing samples' generated n-graphs. Bergadano's approach was based on fixed text. Gunetti et al. extended this to be applicable to free text and therefore, to enable continuous verification [11]. Additionally, they proposed another distance measure based on absolute times. Regarding the combination of the relative and the absolute time approach they achieved better results (FAR < 0.005%, FRR < 5%).

C. *Mouse Dynamics*

Exploiting mouse activities for user verification is a relatively new approach. The pioneering and most comprehensive study to our knowledge was carried out by Ahmed et al [2]. The study defines four different mouse actions as follows: *mouse movement*, *drag and drop*, *point and click* and *silence*. Several different features were defined, such as the interpolation between the movement speed and the traveled distance, which estimates the average speed a user will travel for a certain distance. In addition, several histograms were used to capture different working statistics of the user such as the average travelling speed in eight direction zones or the relative occurrence of each one action.

This study showed relatively good results of less than 3.29% FRR and less than 0.5% FAR, when the number of actions was greater than 2,000 and the verification session last for 13.55 minutes on average. Nevertheless it showed relatively poor results of less than 24% FRR and 4.6% FAR when the session was of a shorter duration (above 4 minutes). The period for identifying the user in this work is far beyond the reasonable time required for an attacker to take full control of a computer system; histograms may reflect different working characteristics of the user but in order for these to be accurate a relatively long time is required, during which an imposter can perform already his malicious act.

Pusara and Brodley [12] attempted to uniquely partition users according to their mouse movement behavior. They calculated the mean, standard deviation and the third moment of the *distance*, *angle* and *speed* between different two adjacent points, when a defined window of data points is considered. A decision tree classifier was trained to differentiate among users activity. Gamboa and Fred [13,14] consider features such as the *angle*, *curvature*, *horizontal*, *vertical* and *combined velocity*, *acceleration* and *jerk* obtained from a vector of data points that were intercepted between two mouse clicks in a web memory game. The authors evaluated the use of two statistical models with the use of the extracted features to verify the identity of an individual.

IV. General Frameworks for Behavioral Biometrics in Computers

In this section we present a general framework, as shown in figure 2, for behavioral biometrics in computers.

A general framework for behavioral biometrics includes several components:

- *Feature Acquisition* – captures the events generated by the varying devices used for the interaction (e.g., keyboard, mouse)
- *Feature Extraction* – extracts a vector of features which describes the biometrics properties of the user.
- *Classifier* – Consists on an *inducer*, commonly is implemented by using classification algorithms (e.g, Support Vector Machines, Artificial Neural Networks and more), which learn the user verification model based on it's past behavior, often given by samples. Later the induced model is used to classify new samples for verification.
- *Signature database* – A database of signatures, which are actually behavioral signatures that were learned by the inducer. Upon a username the signature of the user is retrieved for the verification process.

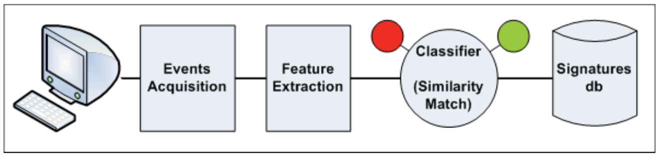

Fig. 2 – A general framework for behavioral biometrics

A. *Main Deployment Configuration*

The two types of methods: login and continuous verification, in which keystroke and mouse dynamics can be exploited to enable an additional security later, can be applied in three types of environments:

- A Host (e.g., a desktop, laptop, or a server), in which

[2] http://psylock.com/index.php/lang-en

all the components of the verification process exist and no communication with the external world is required. In this environment, all the events are being hooked at the operating system kernel level.

- A Web browser in which various web technologies (such as AJAX or Flash) are employed for features acquisitions and all the remaining tasks are performed on the server. The activities of the keystrokes and the mouse that can be accessed are only the ones which are related to the web browser.
- A Client-Server (e.g., a computer within a network of users), in which part of the components are on the client and part on the server side, as we will elaborate in the next sub section B). Here all the events are hooked at the operating system level and sent to the server.

B. *Client-Server Deployment Alternatives*

When using behavioral biometrics for users' verification (log-in or continuous) in Client-Server architecture, there are several deployments, which differ by the computation effort required from the client. This can include only the acquisition of the features and the transmission of the raw data to the server, in which the rest of the computation is being made. Another option is to make the client thicker and containing more functionality. Figure 3 presents three deployments of a behavioral biometrics, as presented in figure 2, in a client-server framework.

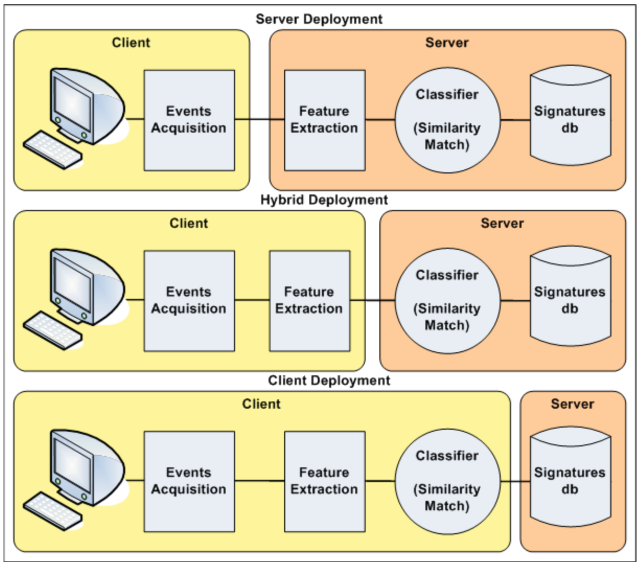

Fig. 3 – Various client-server deployments of the behavioral biometrics framework.

- *Server Deployment* – In the server deployment the only part happening in the client is the collection of events (from the input devices, e.g. keyboard, mouse). These are sent to the server, in which all the rest of the process is performed (feature extraction and similarity match).
  Disadvantages: (1) Sending the events requires a lot of communication. (2) The raw events include private information, which have to be handled carefully at the server.
  Advantage: Easy to update the method, which requires mainly updating the server.
- *Hybrid Deployment* – In the hybrid deployment the client extracts the features from the raw events and sends them to the server, where the matching process is performed.
  Disadvantages: Increased computation overhead for extracting features in the client.
  Advantage: Reduced communication overhead compared to Server Deployment.
- *Client Deployment* – In the Client deployment the signature of the user is downloaded on the first connection, and then all the verification process is performed entirely on the client.

V. Challenges with Behavioral Biometrics in Computers

Having a great potential to overcome the identity theft, through a usable and cheap technology, which we presented here by relying on stroke and mouse dynamics, introduces several challenges towards being operational. In this section we review carefully all the aspects and issues which should be addressed in future work, as well as directions for solutions

A. *Data Collection*

We already mentioned in the literature of Section IV the lack of available data sets for research and evaluation of the proposed methods. Thus, in addition to the lack of a possibility to compare the existing methods (since they create their own dataset, having its own characteristics), each study has to start by putting new efforts in the creation of a dataset.

Datasets, in general, can be of two types: General activities of a user in an operating system of a local computer, in which all the events are hooked at the operating system level, or for a web application, in which all the events, which are related to the web browser, are monitored at the client and sent to the server. The technological aspect of such collection tools is not the problem, but rather the ways to collect authentic data, in which the user performs his daily tasks, as well as attracting many such users.

For the log-in verification one would like to have many users to enter a given password to a website, which are stored in a database. The problem here is mainly to attract users to put the time and the efforts.

Creating a dataset for the *continuous verification* task is more challenging, since the application should be one that demonstrates a real need for verification, and where the users perform their daily tasks and not a specific task given for the experiment. For this purpose, we propose to implement a web-mail interface, in which the users can log into their *real* *account* and work on their *real emails*, but their activities will be monitored and logged at the server. Such requirement from users is reasonable, since it just requires them to do tasks which they have to do anyways. However, the challenge here is the privacy of the users. Thus, it is important not to save the login details and to make sure that the contents of the emails are not exposed. Finally, it is important to make sure that the users are aware of the fact that their activities are logged, in order to avoid future misunderstandings and complaints.

B. Varying Types of Hardware

An important and challenging obstacle of the use of keystroke and mouse dynamics for user verification is that users, especially for web applications, tend to interact from different locations and machines (e.g., their own desktop, laptop or an internet café, etc.). These different machines might include varying types of input devices, such as keyboard, track-point and mouse, and even different machine configurations, which are expected to influence the verification accuracy. In order to investigate these influences, a controlled experiment should be performed in which the users are asked to carry out specific tasks on different computer systems. However, a more authentic approach can be made by asking the users to connect to the application, such as the web-mail interface, from different computer systems.

C. Varying User States

Another influential aspect on the verification accuracy can be caused by different states of the user during the day which might influence their behavior biometrics which is expected to decrease the verification accuracy. The behavior of a user might change along the day, for example in the morning the user is expected to be faster, while slower at night, or after lunch. His physiological position, such as sitting (common), standing, talking on the phone while interacting with the computer, etc. is expected to influence the verification accuracy as well. This of course should be evaluated in a controlled experiment, in which the user is asked to perform the same tasks when being in different states. Since behavior biometrics of a user could change in the long-term use of an application (due to experience), the learning process of the system should be adaptive.

D. Privacy

A major challenge in the use of behavioral biometrics for user verification is the privacy of the users. The "signatures" of the user in terms of keystroke and mouse dynamics are private, as well as the "contents" of the interaction (which is more relevant to typing events). This is a problematic aspect of the dataset collection phase, but also in the operation mode it has to be considered. In the dataset collection it is important to inform the users about the use of the data, and to obtain their prior agreement. Another important issue is the contents of the keystrokes which might contain sensitive and private information such as passwords, which the user has to be aware of and should be able to filter out from the logged information.

E. Scalability

Finally, a very important aspect of this technology is the scalability. Most of the reported papers refer to relatively small evaluation datasets, which might not reflect the challenges that might appear in a real application, where millions of users are present.

VI. Behavioral Biometrics in Mobile Phones – Thoughts of From the Future

In the last decade with the increasing use of mobile phones devices and the increase in their functionalities and storage capabilities, their security becomes crucial.

Clarke and Furnell [15,16,17] studied user verification based on keystroke dynamics in mobile devices, consisting on the keystroke of 11-digit telephone numbers experimenting on text messages and 4-digit PINs to classify users, using feed forward multi-layer perceptron, radial basis function networks, and generalized regression neural networks. Recently, Hwang et al [18] presented an approach for verification of users in mobile devices consisting on 4-digit PIN. They built a classifier using impostors' patterns as well as the valid user's patterns. They proposed adding artificial rhythms to improve the verification accuracy.

However, the recent smart phones, include full keyboards which enable richer activities, such as browsing the websites, as well as touch-phones which enable more opportunities to verify users based on their activates. A major challenge in mobile devices is the current low computing power which should be considered. Exploring these new opportunities were not explored yet to our knowledge.

VII. Discussion

We presented the problem of Identity Theft, in which user's information such as username and password are stolen, which becomes a costly problem. Verifying users based on their keystroke dynamics and mouse activities present a great potential, as an additional biometric security layer. A major advantage of this approach is that it doesn't require additional biometric hardware, which is often costly and not usable, especially in mobile devices.

We presented a general framework for users verification based on their typing and mouse behavior, and several deployments in client-server setup. While several studies were performed in keystroke dynamics in computers during the past decades, the use of mouse activities in computers and the use of mobile phones keystroke were recently being explored. The current methods in mouse activities rely on histograms of activities, which require relatively long periods of time, which enables enough time for an imposter to perform a malicious act and thus approaches, in which significantly faster response time can be accomplished.

Keystroke and Mouse activity based verification was mainly studied for login verification, while continuous verification was not studied almost at all. Continuous verification is crucial to reduce the exploit of logged device, which is stolen or lost. This field requires fast and efficient methods to verify users based on their keystroke dynamics and mouse/touch activities.

Recently the use of behavioral biometrics was proposed for mobile phones in 11-digit keypads, however, the recent growth and acceptance of the smart-phones, which are equipped with full hardware or touch based keyboards and touch-pads, present a richer environment for activity based verification of users.

A major challenge which is open for explorations is the variability in hardware devices, such as varying types of keyboards and mouse devices, as well as the new developments in mobile devices. Additionally, the changing behavior of users along the day is challenging as well. However, overcoming these challenges promise a usable approach for an additional layer of security to overcome the problem of Identity Theft.

ACKNOWLEDGMENT

We would like to thank all the pioneers mentioned in this paper for their contribution in the field, as well as Deutsche Telekom AG for providing funding for the mentioned investigations.

REFERENCES

[1] A Peacock, X Ke, M Wilkerson, Typing Patterns: A Key to User Identification, IEEE Security & Privacy, 1540-7993/04, 2004.

[2] A.A.E. Ahmed, I. Traore, A New Biometric Technology Based on Mouse Dynamics, IEEE Transactions on Dependable and Secure Computing, Vol. 4, No. 3, July-September 2007, pp 165-179.

[3] R. V. Yampolskiy, V. Govindaraju, Behavioral Biometrics: a Survey and Classification, Int. J. Biometrics, Vol. 1, No. 1, 2008.

[4] S. Cho, C. Han, D.H. Han, and H.I. Kim. Web-Based Keystroke Dynamics Identity Verification Using Neural Network. *Journal of Organizational Computing and Electronic Commerce*, 10(4):295–307, 2000.

[5] Daw-Tung Lin. Computer-access authentication with neural network based keystroke identity verification. *Neural Networks,*1997

[6] E. Lau, X. Liu, C. Xiao, and X. Yu. Enhanced User Authentication Through Keystroke Biometrics. Technical report, Massachusetts Institute of Technology, 2004.

[7] K. Revett, P.S. Magalhães, and H.D. Santos. Developing a keystroke dynamics based agent using rough sets. In *2005 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*. University of Technology of Compiègne, 2005.

[8] D. Bartmann and D. Bartmann. Method for verifying the identity of a user of a data processing unit with a keyboard designed to produce alphanumeric characters. European Patent Number EP 0 917 678 B1, 1997.

[9] D. Bartmann and M. Wimmer. Kein Problem mehr mit vergessenen Passwörtern. Datenschutz und Datensicherheit-DuD, 31(3):199–202, 2007

[10] F.Bergadano, D.Gunetti and C.Pcardi, 2002: User Authentication through Keystroke Dynamics. *ACM Transactions on Information and System Security*, Vol 5(4), pp.367-397.

[11] D Gunetti and C Picardi, Keystroke analysis of free text. ACM Trans. Inf. Syst. Secur., 8(3):312–347, 2005.

[12] M. Pusara and C.E. Brodley, User Re-Authentication via Mouse Movements VizSEC/DMSEC'04: Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security, ACM Press, Washington DC, USA, 2004, pp. 1-8.

[13] H. Gamboa and A. Fred, An Identity authentication system based on human computer interaction behaviour, Proc. Of the 3rd IntL. Workshop on Pattern Recognition in Information Systems. ICEIS PRESS, 2003, pp. 46-55.

[14] H. Gamboa and V. A. Fred., A Behavioral Biometric System Based on Human Computer Interaction, In Proceedings of SPIE, 2004.

[15] Clarke N, Furnell S. Authentication of users on mobile telephones – a survey of attitudes and practices. Computers & Security; 24(7):519–27, 2005.

[16] Clarke N, Furnell S. Advanced user authentication for mobile devices. Computers & Security;26(2):109–19, 2007.

[17] Clarke N, Furnell S. Authenticating mobile phone users using keystroke analysis. International Journal of Information Security; 6(1):1–14, 2007.

[18] S Hwang, S Cho, S Park, Keystroke dynamics-based authentication for mobile devices, computers and security, 2 8: 85 – 93, 2009.