ISSN : 2494-9150 Vol-01, Issue 11, FEB 2016.

SECURE SYSTEM USING KEYSTROKE DYNAMICS

[1]Akhilesh R. Vishwakarma, [2]Ganesh Dudhate, [3]Abhishek Bhadale, [4]Prof. Rina Bora
[1,2,3,4]Department of Computer Engineering, Saraswati College of Engineering, Kharghar, Mumbai, Maharashtra, India.
[1]akhi8291@gmail.com, [2]ganeshdudhate@gmail.com, [3]bhadaleak@gmail.com, [4]rkbora2006@gmail.com

Abstract **— Research on keystroke dynamics biometrics has been increasing, especially in the last decade. The main motivation behind this effort is due to the fact that keystroke dynamics biometrics is economical and can be easily integrated into the existing computer security systems with minimal alteration and user intervention. Having a secure information system depends on successful authentication of legitimate users so as to prevent attacks from fraudulent persons. Traditional information security systems use a password or Personal Identification Number (PIN). This means they can be easily accessed by unauthorized persons without access being noticed. Numerous studies have been conducted in terms of data acquisition devices, feature representations, classification methods, experimental protocols, and evaluations. However, an up-to-date extensive survey and evaluation is not yet available. This project addresses the issue of enhancing such systems using keystroke biometrics as a translucent level of user authentication. The project focuses on using the time interval (key down-down) between keystrokes as a feature of individuals' typing patterns to recognize authentic users and reject imposters. A Multilayer Perceptron (MLP) neural network with a Back Propagation (BP) learning algorithm is used to train and validate the features.**

Keywords— keystroke dynamics biometrics, neural keystroke algorithm.

I. INTRODUCTION

The number of computer uses has increased rapidly and so too has the use of internet applications such as e-commerce, online banking services, webmail, and blogs. All internet applications require the user to use a password authentication Scheme to make sure only the genuine individual can login to the application. Passwords and personal identification Numbers (PIN) have traditionally been used to access such applications. However, it is easy for unauthorized persons to access these systems without detection. In order to enhance such password authentication systems, typing biometrics,

Known as keystroke, can be used as a transparent layer of user authentication. Keystroke verification techniques can be categorized as either static or continuous. Static verification system approaches study keystroke characteristics at a specific time. Although they are more robust they cannot detect a substitution of the user after initial verification. Continuous verification, on the other hand, examines the user's typing behavior throughout the interaction time. Time-features can be extracted from keystroke data in many ways, such as studying keystroke latency, duration of key hold, pressure of keystroke, frequency of word errors, and typing rate. However, not all of these methods are widely used. Keystroke solutions are usually measured in three ways: dwell time – how long a key is pressed, flight time – how long it takes to move from one key to another, and key code. Keystroke dynamics is one of the novel and creative biometric techniques. It is not only nonintrusive, but also transparent and inexpensive.

II. LITERATURE REVIEW

The emergence of keystroke dynamics biometrics was dated back in the late 19th century, where telegraph revolution was at its peak [1]. It was the major long distance communication instrument in that era. Telegraph operators could seamlessly distinguish each other by merely listening to the tapping rhythm of dots and dashes. While telegraph key served as an input device in those days, likewise, computer keyboard, mobile keypad, and touch screen are common input devices in the 21st century. Furthermore, it has been noted that keystroke pattern has the same neurophysiologic factors that make hand written signature unique [2], where humans have relied on to verify identity of an individual for many centuries. In fact, keystroke pattern is capable of providing even more unique feature for authentication, which includes key press duration and latencies, typing rate, and typing pressure. Among the earliest significant keystroke dynamics research work on authentication was conducted by [3], ever since, this domain has gradually gained momentum (Figure 1). Figure 2 shows the timeline development in the area of keystroke dynamics biometrics.

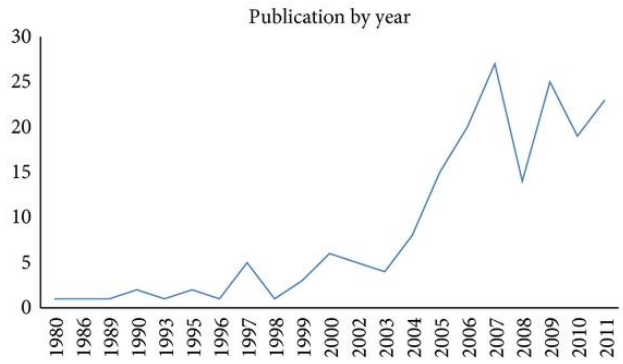

Figure 1 Graph clearly indicates an increasing trend on research work conducted on keystroke dynamics domain.

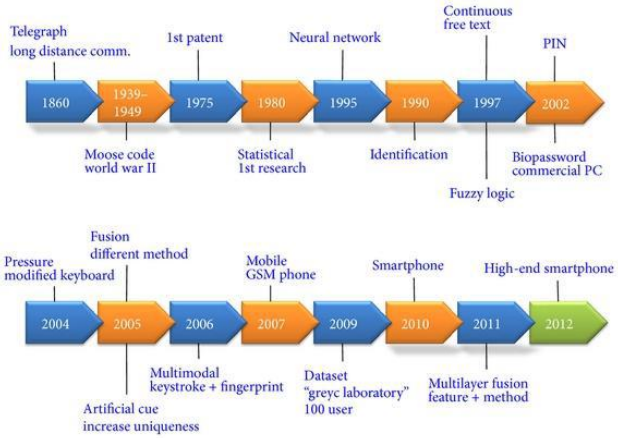

Figure 2 A general timeline on the overview of keystroke research work evolution.

A) Feature Selection

Keystroke dynamics biometrics are rich with distinctive feature information that can be used for recognition purposes. Among the easiest and common feature harvested by researchers is the timing measurement of individuals' keystroke inputs as shown in Figure 3.

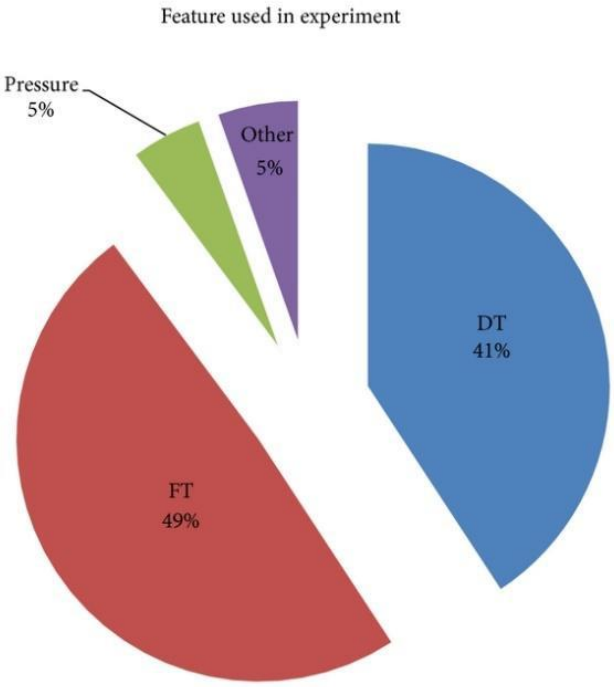

Figure 3 thepercentage distribution of feature data extracted for keystroke experiment in the literature.

Keystroke activity generates hardware interrupt that can be time stamped and measured up to microseconds (ms)

precision [4]; therefore, it can be readily applied. In previous works, timing resolution of 0.1 s to 1ms has been deemed to be sufficient [5]. By performing simple mathematical operation to these time stamp, timing duration, or interval between consecutive keystrokes can be obtained.

Several attempts, although uncommon, of using keystroke pressure, typing speed , typing sequence difficulty, frequency of typing error , and sound of typing have also been made. Due to the insignificant amount and unpopularity of the aforementioned feature type, the following subsections will focus on the discussion of the more popular timing feature.

Timing information of two consecutive keystrokes, better known as di-graph, is the major feature data represented in keystroke dynamics domain. It is widely categorized into two types, namely, Dwell Time and Flight Time. Both are relatively equally weighted in terms of usage frequency among 187 research works as illustrated in Figure 3.

B) Dwell Time (DT)

Dwell time refers to the amount of time between pressing and releasing a single key. In other words, how long a key was held pressing down. It is also worth noticing that several terms for DT appeared in the literature such as duration time [] and hold time. DT can be calculated by

$$DTn = Rn - Pn$$

Where RP indicate the time stamp of release and press of a character, respectively, while n indicates the position of the intended DT.

For instance, referring to Figure DT for character “J” and “Y” is 100 (200–100) and 250 (750–500) correspondingly. The total number of timing vector of DT (VDT) that can be generated as follow:

$$VDT = \{DT1, DT2, DT3\ldots DTS\},$$

s denotes the summation of characters in a string. In other words, the number of DT generated will always be the same as the length of a given string.

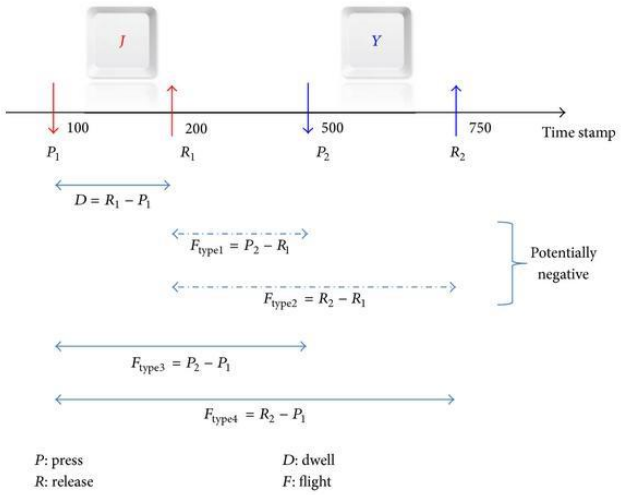

Figure 4: Figure depicts the different keystroke events of two characters “J” and “Y” alongside with the formation of dwell time and flight time.

C) Flight Time (FT)

Flight time refers to the amount of time between pressing and releasing two successive keys. It may also be termed as latency time, inter-key time or interval time. It always involves key event (press or release) from two keys, which could be similar or different characters. FT may exist in four different forms as depicted in Figure 4. The formula to calculate each form are listed as follows:

$$FTtype1,n = Pn+1 – Rn$$

$$FTtype2,n = Rn+1 – Rn$$

$$FTtype3,n = Pn+1 – Pn$$

$$FTtype4, n = Rn+1 – Pn$$

RPn indicates the position of the intended FT.

As an example FTtype1 between character “JY” shown in Figure is 300 (500–200), whereas the FTtype3 is 400 (500–100). The previous literature pointed out the possibility of obtaining negative value (<0). This situation occurs when an individual presses the next key before releasing the previous key. However, a closer observation shows that it is also possible for FTtype2 to incur this property, albeit in a very

exceptional circumstance. The total number of timing vector of FT (VFT) that can be generated is shown as follows:

VFT = {FT1, FT2, FT3… FTs-1}

Denotes the summation of characters in a string. Differing from, the number of generated will always be one less than the length of a given string.

III. EXISTING SYSTEM

Authentication is an important factor in computer and network security. Whether it is controlling access to company resources or verifying the billing of a customer on an online shopping website, verifying a user's identity will always be an integral part of a secure system. There are many techniques currently used in authentication. The most common is the password. Passwords are convenient as they are easily implemented in software and require no specialized hardware. Users are also familiar with their use. Passwords also suffer from many flaws. Users frequently share passwords, forget passwords, and select poor passwords that may be easily defeated. This has spawned research and innovation into alternatives to supplement and supplant the common password.

Password Alternatives

Physical security measures, such as access cards or keys, are one such alternative. They are reliable provided the physical devices are not lost or stolen. Frequently they are paired with simple passwords to reduce the loss of security with theft. For example the PIN number associated with automatic teller machines. Physical security techniques are useful but rely heavily on expensive customized hardware and software products.

IV. SCOPE OF THE PROJECT

Technology development over the past decade has contributed to the escalating access and storage of confidential information in digital devices. Therefore, the need for a more secure authentication mechanism becomes imminent. Authentication in short is the process of verifying a person's legitimate right prior to the release of secure resources. Generally this is achieved by counterchecking unique information provided by an individual. This information can be broadly subdivided into three categories namely knowledge, token, and biometrics-based authentication as summarized in Table 1 and discussed as follow.

Approach	Advantage	Disadvantage	Example
Knowledge	Effortless High acceptance	Forgotten Shoulder spoofing	Password PIN
Token	Cheap Simple deployment	Lost and theft	Smart card Minidevices
Biometrics	Deter sharing Unique Unforgettable	Cost Invasive	Fingerprint Voice Keystroke

Table 1: Overview of different authentication approaches.

A) Knowledge

Knowledge commonly regard as something a person knows [6], which generally resides in the form of texture or graphical password, personal identification number (PIN), and pattern code. Password-based authentication has been an established method for access control in variety of systems since the past three decades [7]. Cost effectiveness and simple implementation have been the forefront reasons for the continuous dominance of password. Nevertheless, the ability for it to provide confident and secure authentication has been wearing, due to reasons such as the wrongful use of password and increased intrusion attacks. Simple password is the primary choice when it comes to password selection, such as date of birth, nickname, initials, and regular dictionary words that is either easily guessed or hacked. To aggravate the situation, users always tend to use the same or similar password for multiple systems. These

bad usage habits contribute to the deterioration of knowledge-based authentication quality.

B) Token

Token refers to an object that requires user to physically possess as a form of authentication. Common tokens include but not limited to swipe cards, credit cards, and mini devices. Although large-scale deployment is relatively simple [8], it comes with its own weakness. Token are vulnerable to loss or theft as user may find it inconvenient or difficult to keep it safe at all times. This implies that there is no assurance on uniquely identifying a legitimate user even with the ownership of token. Typically this shortcoming can be resolved by using token alongside knowledge-based method. At such, these two entities together render a simple two-factor authentication process that produces a stronger authentication based on the assumption that the secrecy of knowledge is not breached.

C) Biometrics

Biometrics refers to certain physiological or behavioural characteristic that is uniquely associated to a person. This trait is highly distinctive and can be utilized for distinguishing different individuals. Physiological biometrics refers to a person's physical attribute, such as fingerprint, face, and iris. It is well known for its permanence and high uniqueness that promote high recognition accuracy. Unfortunately, it is not likely to be revoked if compromised (unable to change fingerprint pattern) [9], may possibly suffer low public acceptance due to invasiveness (iris scanning), and could be unlikely practical in large-scale deployment due to implementation cost (DNA analysis).The way people do things such as speaking (voice), writing (signature), typing (keystroke dynamics), and walking style (gait recognition) are known as behavioural biometrics. Behavioural biometrics has the edge over its physiological counterpart on the ability to work in stealth mode verification. As such, minimal interaction is required during authentication process reduces invasiveness and thus promotes user acceptability. In addition, in the event if one's behavioural attribute is compromised, it is likely to be replaced (changing to a new password, thus, new keystroke print or new written signature) [10]. While these merits may be encouraging, they are normally inferior to physiological biometrics in terms of variability (voice changes along with aging factor) and may consequently influence verification accuracy.

V. SYSTEM REQUIREMENT SPECIFICATION

1. Project Language

1.1. **Front End:** Microsoft C#.Net 2008

1.2. **Back End:** Microsoft SQL Server 2005

2. About Visual Studio

The Microsoft Visual Studio development system is a suite of development tools designed to aid software developers—whether they are novices or seasoned professionals—face complex challenges and create innovative solutions. Every day, software developers break through tough problems to create software that makes a difference in the lives of others. Visual Studio's role is to improve the process of development to make the work of achieving those breakthroughs easier and more satisfying. How Visual Studio improves the process of development:

3. Productive

Visual Studio-branded tools continually deliver better ways for software developers to do more with less energy wasted on repetition and drudgery. From efficient code editors, IntelliSense, Wizards, and multiple coding languages in one integrated development environment (IDE) to high-end application life-cycle management (ALM) products in Microsoft® Visual Studio® Team System. New versions of Visual Studio keep bringing innovative tools to help developers focus on solving problems, not waste time on minutiae.

4. Integrated

With Visual Studio, software developers benefit from an integrated product experience that spans tools, servers, and services. Visual Studio products work well together—not just with one another, but also with other Microsoft software, such as Microsoft server products and the Microsoft Office system.

5. Reliable

Visual Studio is engineered and tested to be consistently dependable, secure, interoperable, and compatible. Visual Studio offers an unmatched combination of security features, scalability, and interoperability. Although Visual Studio always incorporates forward-thinking features, it is designed to ensure backward-compatibility wherever possible.

6. Visual Studio and the Microsoft Application Platform

The Microsoft Application Platform is a portfolio of technology capabilities, core products, and best practice guidance focused on helping IT and development department's partner with the business to maximize opportunity.

As one of the core products of the Microsoft Application Platform, Visual Studio can help you drive the right business efficiencies, customer connections, and value-added services by providing a single, fully integrated development environment for all types of development, including Microsoft Windows, Microsoft Office, Web, and mobile applications. Use Visual Studio development solutions to give your development team powerful ways to:

- Increase productivity and quality through integrated and familiar tools.
- Deploy, secure, and support your critical Web applications and infrastructure.
- Reduce costs through better visibility of your development process.
- Provide better predictability and planning through integrated process and methodology support.

7. About SQL Server

Microsoft SQL Server 2005 is a comprehensive, integrated data management and analysis software that enables organizations to reliably manage mission-critical information and confidently run today's increasingly complex business applications. SQL Server 2005 allows companies to gain greater insight from their business information and achieve faster results for a competitive advantage.

VI. SYSTEM ARCHITECTURE

It has been shown that a layered neural network provides more potential alternatives than traditional pattern recognition techniques (Burr, 1988; Anagun and Liou, 1993). The neural networks used in this study, one for each type of data, were made up of three layers with inter-layer connections. The number of neurons in the network architecture was varied depending on the experiments. The input layer was composed of 8-13 neurons, represented time intervals between successive keystrokes obtained from the passwords entered, and 75-100, represented LPCs obtained from the transformation process for the passwords spoken. For both experiments, the output layer was made up of 3-6 neurons for the desired output values of each pattern; for instance, 100100 represented that the first user entered/spoken the first password. The number of neurons in the hidden layer, which yields to extract features between the input and the corresponding output pattern, was varied depending on the experiments to improve the network performance in terms of generalization. The learning rate and momentum term were arbitrarily assigned to 0.15 and 0.4, respectively.

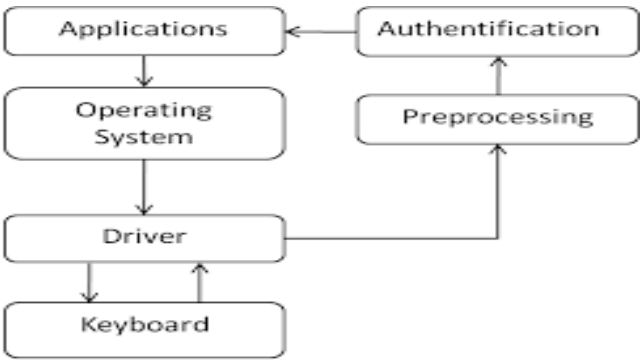

Figure 5. Flowchart of the keystroke dynamics system.

VII. CONCLUSION

Keystroke analysis has proven itself capable of providing additional security above and beyond conventional passwords. The lack of special hardware and resulting low cost make it a popular area in biometrics research. The literature study suggests that keystroke dynamics biometrics are unlikely to replace existing knowledge-based authentication entirely and it is also not robust enough to be a sole biometric authenticator. However, the advantage of keystroke dynamics is indisputable such as the ability to operate in stealth mode, low implementation cost, high user acceptance, and ease of integration to existing security systems. These create the basis of a potentially effective way of enhancing overall security rating by playing a significant role in part of a larger multifactor authentication mechanism.

REFERENCES

[1] BioPassword, Authentication Solutions Through Keystroke Dynamics, BioPassword, Issaquah, Wash, USA, 2006.
[2] A. K. Jain, R. Bolle, S. Pankanti, M. S. Obaidat, and B. Sadoun, "Keystroke dynamics based authentication," in Biometrics, pp. 213–229, Springer, New York, NY, USA, 2002.
[3] R. S. Gaines, W. Lisowski, S. J. Press, and N. Shapiro, "Authentication by keystroke timing: some preliminary results," Tech. Rep. R-2526-NSF, Rand Corporation, Santa Monica, Calif, USA, 1980.
[4] S. J. Shepherd, "Continuous authentication by analysis of keyboard typing characteristics," inProceedings of the 1995 European Convention on Security and Detection, pp. 111–114, May 1995. View at Scopus.
[5] D. Hosseinzadeh and S. Krishnan, "Gaussian mixture modeling of keystroke patterns for biometric applications," IEEE Transactions on Systems, Man and Cybernetics C, vol. 38, no. 6, pp. 816–826, 2008.View at Publisher · View at Google Scholar · View at Scopus .
[6] S. J. Shepherd, "Continuous authentication by analysis of keyboard typing characteristics," inProceedings of the 1995 European Convention on Security and Detection, pp. 111–114, May 1995.
[7] M. Karnan and M. Akila, "Identity authentication based on keystroke dynamics using genetic algorithm and particle swarm optimization," in Proceedings of the 2nd IEEE International Conference on Computer Science and Information Technology (ICCSIT '09), pp. 203–207, August 2009.
[8] P. S. Teh, A. B. J. Teoh, C. Tee, and T. S. Ong, "A multiple layer fusion approach on keystroke dynamics," Pattern Analysis and Applications, vol. 14, no. 1, pp. 23–36, 2011.
[9] B. Ngugi, B. K. Kahn, and M. Tremaine, "Typing biometrics: impact of human learning on performance quality," Journal of Data and Information Quality, vol. 2, no. 2, article 11, 2011.
[10] B. Ngugi, M. Tremaine, and P. Tarasewich, "Biometric keypads: improving accuracy through optimal PIN selection," Decision Support Systems, vol. 50, no. 4, pp. 769–776, 2011.
[11] Performance of Keystroke Biometrics Authentication System Using Artificial Neural Network (ANN) and Distance Classifier Method, By N. Harun, W. L. Woo and S.S. Dlay; Electrical, Electronic and Computer Engineering, Newcastle University, United Kingdom.

