Security Enhancements for Person Verification Using Multi Model Fusion Keystroke Dynamics and Soft Biometrics

by

Mohd Noorulfakhri bin Yaacob
(1643112272)

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

Faculty of Applied and Human Sciences
UNIVERSITI MALAYSIA PERLIS

2021

ACKNOWLEDGEMENT

Alhamdulillah, praise and thanks to Allah 'Azza wa Jalla for His abundant grace I was able to complete my PhD study and fulfilled the requirements at the Faculty of Applied and Human Sciences (FSGM), Universiti Malaysia Perlis (UniMAP).

I am grateful that all of the predicaments and obstacles encountered during this study were successfully overcome.

Many thanks to Associate Professor Ts. Dr. Syed Zulkarnain bin Syed Idrus Al-Saggoff (Lead Supervisor) and Prof. Dr. Christophe Rosenberger from University of Normandy, France (Co-Supervisor) for their assistances in giving me vital advices during the course of my study. Their dedications in giving me guidance, encouragements and advices will be my lifelong memory.

An unpaid debt of gratitude to Universiti Malaysia Perlis as my employer that had given me their full support by reducing my education costs.

Finally, I would also like to take this opportunity to thank all my family members and friends for their constructive encouragements during the difficult times and crucial situations that I had on completing this study. They are my parents, my wife Siti Masliana binti Bohari, my daughters and all of my good friends who had helped make this study possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AaMLP	Auto Associative Multilayer Perceptron
ANN	Artificial neural networks
ATM	automated teller machine
BMI	Body mass index
BPPN	Back Propagation Neural Network
BRF	Balanced Random Forest
CER	control event rate
CGPA	Cumulative Grade Point Average
CMU	Carnegie Mellon University
CPU	Central Processing Unit
DA-SVM	Deformable Adaptive
DNA	Deoxyribonucleic Acid
DNN	Deep Neural Networks
DT	Dwelling Time
EEG	Electroencephalogram
EER	Equal Error Rate
FANN	Fast Artificial Neural Network
FAR	False Acceptance Rate
FNMR	False Non-Match Rate
FMR	False Match Rate
FRNN-VQRS	Fuzzy-Rough Nearest Neighbor Vaguely Quantified Rough Set
FRR	False Rejection Rate
FT	Flight Time
GREYC	Groupe De Recherche En Informatique, Image, Automatique Et Instrumentation De Caen
GMM	Gaussian Mixture Model
HMOG	Hand Movement, Orientation, And Grasp
HSDPA	High-Speed Downlink Packet Access
IBM	International Business Machines
ID	Identification
IT	Information Technology
ICT	Information and Communication Technology
MB	Megabyte
MR	Malaysia Race

MFN	Multilayer Feedforward Network
MLP	Multilayer Perceptron
PMT-SVM	Projective Model Transfer
PNN	Probabilistic Neural Network
PP	Press-Press
PR	Press-Release
PUC	Pairwise User Coupling
RAM	Random-Access Memory
RBF	Radial Base Function
RBFN	Radial Basic Function Network
RF	Random Forest
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
ROS	Random Over Sampling
RP	Release-Press
RR	Release-Release
RUS	Random Under Sampling
SMD	Scaled Manhattan Distance
SMOTE	Synthetic Minority Oversampling Technique
STD	Standard Deviation
SVM	Support Vector Machine
UED	Using End-User Development
UKKL	Up-Up Keystroke Latency
USB	Universal Serial Bus
VCS	Verification Combine Score

LIST OF SYMBOLS

SBc	Soft Biometric Class
SBs	Soft Biometric Score
SBv	Soft Biometric Value
σ	Standard Deviation
μ	Mean
VCS	Verification Combine Score
Ds	Distance Score
SBz	Total Soft Biometric Value

Pemantapan Keselamatan Untuk Pengecaman Manusia Menggunakan Gabungan Pelbagai Model Bagi *Keystroke Dynamics* Dan Biometrik Lembut

ABSTRAK

Biometrik digunakan sebagai pagar keselamatan utama di dalam sistem komputer. Biometrik yang terdapat pada diri manusia ini boleh dikategorikan kepada tiga jenis iaitu biologikal, tingkah laku dan morfologi. Kajian ini memberi fokus kepada biometrik tingkah laku iaitu dengan mengkaji berkaitan dinamik ketukan kekunci (*Keystroke dynamics)*. *Keystroke dynamics* digunakan sebagai pengesahan peringkat kedua untuk keselamatan sistem yang menggunakan papan kekunci sebagai kawalan akses. Biometrik lembut pula adalah satu kaedah di mana kita dapat mengenalpasti identiti manusia berdasarkan ciri-ciri fizikal atau tingkah laku seseorang yang terhasil secara semulajadi. Masalah yang biasa dihadapi dalam pencerobohan sistem adalah apabila sistem gagal mengenalpasti pengguna sebenar yang masuk ke sistem setelah menggunakan kata kunci yang tepat. Maklumat seseorang pengguna boleh dicuri, diteka dan direkodkan oleh perisian khas menggunakan *keylogger*. Berkemungkinan juga terdapat pengguna yang mampu menceroboh ke dalam sistem dengan menggunakan kaedah serangan *brute-force* and serangan kamus *(dictionary attacks)*. Bagi memastikan dan meningkatkan proses pengesahan pengguna yang menggunakan papan kekunci sebagai kaedah memasuki sistem, kajian ini menggunakan kaedah gabungan pelbagai model bersama *Keystroke dynamics* dan biometrik lembut. Kajian ini memperkenalkan beberapa elemen-elemen biometrik lembut iaitu kaum, tahap pengajian, kawasan dan jantina di dalam *Keystroke dynamics*. *Keystroke dynamics* digunakan bagi membezakan corak menaip bagi setiap kategori ini dan seterusnya ciri-ciri biometrik lembut ini digunakan untuk meningkatkan lagi proses pengesahan pengguna. Formula matematik seperti purata, sisihan piawai dan *Scale Manhattan Distance* telah digunakan untuk mencapai nilai skor penyamar dan pengguna yang sah. Kaedah gabungan pelbagai model telah diperkenalkan dengan mengintegrasikan nilai skor beberapa elemen biometrik lembut dari klasifikasi dengan skor yang diperolehi dari penyamar dan pengguna sebenar bagi menambahbaik proses pengesahan. Bagi tujuan identifikasi, kaedah pengkelasan Mesin Sokongan Vektor digunakan untuk pengkelasan biometrik lembut. Hasil ujian yang diperolehi menunjukkan perbezaan yang ketara pada corak menaip bagi kategori kawasan. Ketepatan tertinggi yang diperolehi adalah 91.17% bagi kategori kawasan. Kawasan ralat yang sama yang diperolehi dari kaedah gabungan pelbagai model adalah 11.33% dengan menggabungkan empat ciri biometrik lembut iaitu jantina, kaum, kawasan dan tahap pendidikan di dalam proses pengesahan. Secara keseluruhannnya didapati proses gabungan pelbagai model yang diperkenalkan ini telah membantu mengurangkan nilai kawasan ralat yang sama bagi tujuan pengesahan pengguna di dalam *Keystroke dynamics*.

Security Enhancements for Person Verification Using Multi Model Fusion Keystroke Dynamics and Soft Biometrics

ABSTRACT

Biometric is used as a main security fence in a computer system. Human's biometrics can be categorized into three types: morphological, biological and behavioral. This study focuses on behavioral biometric by studying keystroke dynamics. Keystroke dynamics is used as the user's second level verification for the systems that use the keyboard to login into a system. Soft biometric is a method by which we can identify human identity based on the physical characteristics or behaviors of a person that are naturally produced. A common problem of system intrusions is that the system fails to identify the real user who signs in using the keyboard when the login is correct. A user's login information can be stolen, guessed, and recorded by special software such as a keylogger. There is a possibility that someone else tries to break into the system by using brute-force and dictionary attacks. To ensure and improve user's verification who use the keyboard to enter their logins into the system, this study has used multi model fusion methods to combine keystroke dynamics and soft biometric. This study introduces new soft biometric elements in keystroke dynamics that is gender, race, region and educational level. Keystroke dynamics is used to distinguish typing patterns in each of these categories and uses these soft biometric features to further enhance the verification capabilities. Mathematical formulas such as mean, standard deviation and Scale Manhattan Distance have been used to obtain the score of the imposter and genuine user. Multi model fusion methods have been introduced by integrating the score values of several soft biometric elements from classification with scores obtained from imposter and genuine user to improve the verification process. For identification purposes, Support Vector Machine classification method is used to perform this classification for soft biometric identification. The results show that there are significant differences in the typing pattern in the region category. The highest accuracy achieved is 91.17% in this classification process for the region category. Equal error rate (EER) obtained from a multi model fusion approach is 11.33%. which is to incorporate four features of soft biometrics (gender, race, region and educational level) into the verification process. Overall, it is found that the multi fusion process introduced has managed to reduce the EER values for the purpose of verification in keystroke dynamics.

CHAPTER 1 : INTRODUCTION

1.1 Background of Study

The developments in the field of Information and Communication Technology (ICT) have begun drastically during the mid 1990s since the introduction of the Internet worldwide. At this present day, almost every system that ever developed would require an Internet connection as the core of their systems (Sunyaev, 2020). With the existence of Internet technology, users are now able to access information they need immediately and easily using personal computers or mobile phone via latest broadband technology such as 3G, Streamyx, Unifi and High-Speed Downlink Packet Access (HSDPA).

Alongside with the advancement of the Internet, developed nations are also associated with the advanced computing systems such as e-governments, e-commerce, and e-medicine. These systems are meant to execute critical tasks in managing data, which mostly are confidential residing in the accounting information systems, military control systems and patient information systems (Gupta et al., 2020; Patnala et al., 2020). Most systems with these confidential information have certain security features to secure data. Nonetheless, user verification is the main security measure deployed for each system to safeguard access to any systems. This shows the importance of computer security in the process to secure vital information.

Computer security also known as Information Technology (IT) security or cyber security includes protection and defense of important information in a system against

being hacked or exposed to users without authorities to that information (Gasser, 1988). Thus, computer security can be categorized into three main parts:

i Hardware Security

Hardware security protects against any potential security risk and vulnerability using the hardware being used by an application system. For example, security of a storage device. Every data stored in storage devices must be encrypted (Fadiheh et al., 2019; Riazi et al., 2020).

ii Network Security

Network Security is the process of taking physical and software preventative measures to protect the underlying networking infrastructure from unauthorized access, misuse, malfunction, modification, destruction, or improper disclosure, thereby creating a secure platform for computers, users and programs to perform their permitted critical functions within a secure environment. (Abdulrahman & Jumaa, 2016)

iii Software Security

Software security is a process to control programming during system development so it would not be easy to hacked into or get attacked by outsiders (van Oorschot, 2020).

Combination of all these three security aspects are very important to secure entire confidential information inside a system. Intruders will always peek upon vulnerability of each aspect before carrying out any attacks. Therefore, thorough and deep research on each aspect is needed to ensure data safety from being hacked or abused. Various hardware have been introduced to help reduce the risk of cyber attacks. Multiple hardware products known to help stop cyber attack are RazorGate, SurfControl, NetEnforcer, Fortinet, Bluecoat and Cisco IronPort Web Security (Edward, 2019). All these devices utilize latest technologies to enable them to operate well.

Despite multiple methods introduced, research and new technologies to secure computers against intrusion, where cases are still being reported until now (Patrick & Fields, 2017). Attack and intrusion on computer systems became the main issue and agenda on data security. Report by Anti-Phishing Working Group shows that as many as 146,994 phishing attacks occurred from April to June 2020 (Aaron, 2020). Phishing is a method employed by cyber criminal to steal personal information using a combination of malware and social engineering. Most current malware designs utilize social engineering methods to deceive users.

Cyber criminals also trap Internet users by stealing their identity such as obtaining their users' identification (ID) from a particular system. They would deceive their victims via fake emails sent to the users. Subsequently, these cyber criminals would use stolen identities to obtain other confidential data and information (Dastbaz et al., 2018). The followings are the latest reports on intrusion incidents globally in general:

i Dropbox - Online storage provider company reported that up to 68 million email and password were stolen by hackers since 2012 (Jones, 2016).

ii Email attacks - News on 27 September 2016 reported that government of United States will be taking actions against cyber intruders trying to access Gmail accounts of the White House personnel (Jones, 2016).

Malaysia is also not spared from cyber threats. According to the reports produced by MyCert, up until December 2019, reports on computer security threats have not been reduced, since the record began in 1997 (CyberSecurity, 2019) . Various computer threats and attacks have been reported and recorded. Figure 1.1 shows the report on threats produced by MyCert as of December 2019, where there are nine categories of cyber threats recorded by the agency. This proves that the research on improving computer security are still needed.

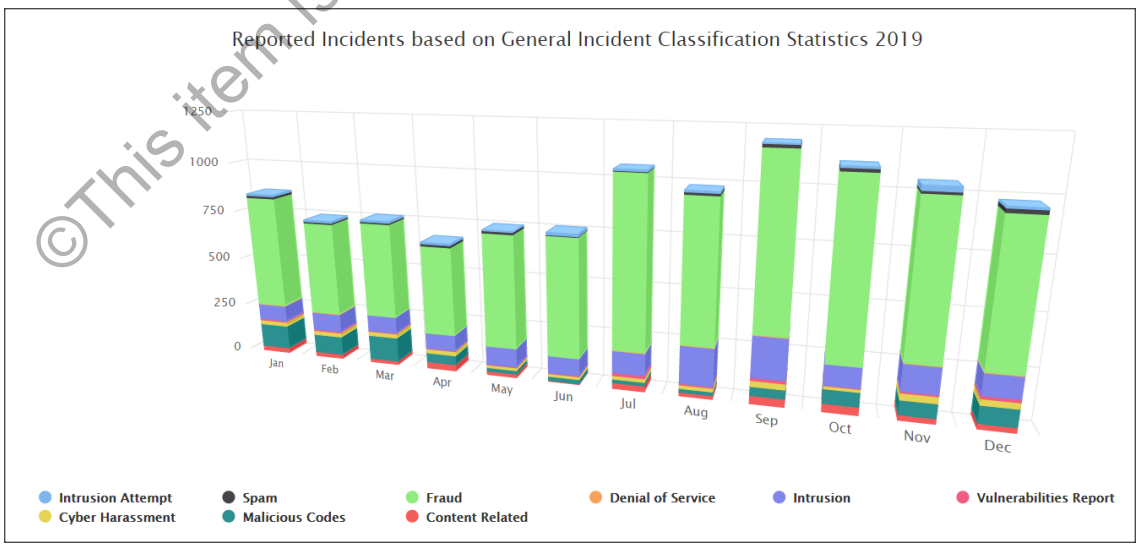

Figure 1.1 Cyber threat reports by MyCert (CyberSecurity, 2019)

In order to handle issues pertaining to computer security, most developed applications are built with a user verification screen to allow the users to input their user IDs and passwords before using the system. This method is one way to secure computer through the software security aspect and also known as user verification.

There are three methods implementable to authenticate users (Monrose, 2000):

i **Something the user knows**. It is related to things that users know such as user ID and password. This method is the most preferred method amongst systems developer to be applied in the system that is being developed. Most systems in use today utilize this method for system usage authorization. In order to use this method, users are required to input one unique user id and password. However, research has found that users regularly used easy or same passwords for all application (Yıldırım & Mackie, 2019). This practice made it easy for user id and password to be guessed using dictionary attack.

ii **Something the user has.** This method uses additional physical accessories owned by registered user. Simplest example for this method is house or office key; only person with the key are allowed access to the place. This method is applied in computer system with additional accessory given to user such as USB key (Gregory, 2015), token (Suwald & Burghardt, 2020) and Smart Card (Hussain et al., 2019). The user of these additional accessories is required to keep them from being lost, stolen or copied by others. Should there be an incident of misplaced, stolen

or unauthorized duplication, users are bound to report to system administrator to block the particular device.

iii **Something the user is.** It is about unique features of users or most commonly known as biometrics. User identification using biometrics methods have started since the last few decades. Biometrics technology is closely related with human physical body or features on human. This method is divided into three main categories: Morphological, Behavioral and Biological (Choudhury, 2020; Tripathi, 2011) (refer to Figure 1.2)).

a) **Morphological** biometrics uses physical features on user for identification. Example for this feature would be fingerprints(Sokolov & Ponomareva, 2020), iris or eye retina (Hájek & Drahanský, 2019), facial shape (Best-Rowden & Jain, 2017), and palm geometry (Alpar & Krejcar, 2017). This method is dependable since it is hard to change its ownership or borrowed (Nawaya et al., 2019).

b) **Behavioral** biometrics is related to the behavior of a system user. Measurement and verification for this method can be either voice identification (Kuśmierczyk et al., 2020), keyboard typing style (Mondal & Bours, 2017), walking style (Condell et al., 2017) and signature (Kancharla et al., 2018). This measurement criteria is done on user spontaneously or without user noticing so that the real user behavior can be recorded (Alrumaih et al., 2020).

c) **Biological** biometrics is a method to identify user based on biological features available on users. Example for user identification using this technology would be DNA, molecule structure, bones, teeth and internal organs (Correa et al., 2020).

Each of this method has its own advantages and disadvantages. The actual usage of physical biometrics, smart card, USB Key and token in a system would involve high costs (A. V. S. Kumar, 2019). Each user wishes to use the system with those features are required to purchase related equipments or devices, at a higher price. On the contrary, methods of using user ID and password, and behavioral biometrics do not require additional cost for equipments (Saeed, 2016). However, only a few methods of biometric user identification have been utilized by commercial systems to perform verification of the validity of users. For example, Morphological identification method using finger print, iris and face recognition have been used by the security force of the United States to control the country border entry point (Buhrow, 2016).

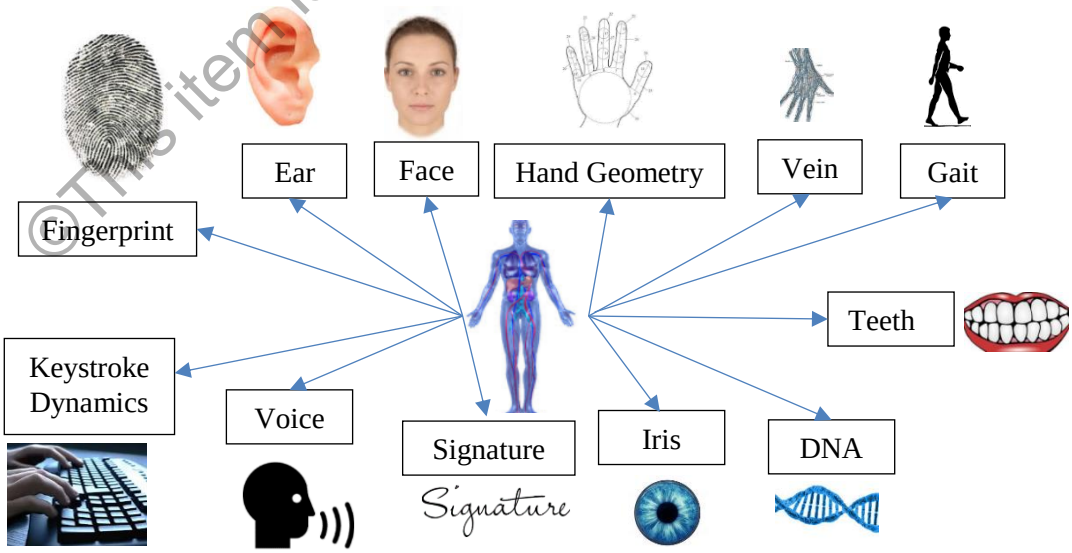

Figure 1.2 Examples of biometric modalities that can be used to authenticate the users