

**DEVELOPMENT OF STATIC KEYSTROKE DYNAMICS
AUTHENTICATION SYSTEM**



UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS LAPORAN

JUDUL: [DEVELOPMENT OF STATIC KEYSTROKE DYNAMICS AUTHENTICATION SYSTEM]

SESI PENGAJIAN: [2022 / 2023]

Saya: NG WAN THONG

mengaku membenarkan tesis Projek Sarjana Muda ini disimpan di Perpustakaan Universiti Teknikal Malaysia Melaka dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. * Sila tandakan (✓)

SULIT

(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

TERHAD

(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi / badan di mana penyelidikan dijalankan)

TIDAK TERHAD


(TANDATANGAN PELAJAR)

Alamat tetap: No 38, Jalan Marin2,
Taman Marin 84000 Muar Johor


(TANDATANGAN PENYELIA)

PROFESOR MADYA DR MOHD FAIZAL ABDOLLAH

Nama Penyelia

Tarikh: 21/9/2023

Tarikh: 22/9/23

CATATAN: * Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak

DEVELOPMENT OF STATIC KEYSTROKE DYNAMICS AUTHENTICATION
SYSTEM

NG WAN THONG



This report is submitted in partial fulfillment of the requirements for the
Bachelor of Computer Science (Computer Security) with Honours.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2023

DECLARATION


I hereby declare that this project report entitled
**DEVELOPMENT OF STATIC KEYSTROKE DYNAMICS AUTHENTICATION
SYSTEM**

is written by me and is my own effort and that no part has been plagiarized
without citations.

STUDENT :  Date : 21/9/2023
(NG WAN THONG)



I hereby declare that I have read this project report and found
this project report is sufficient in term of the scope and quality for the award of
Bachelor of Computer Science (Computer Security) with Honours.

SUPERVISOR :  Date : 22/9/23
(PROF. MADYA. DR. MOHD FAIZAL BIN ABDOLLAH)

DEDICATION

This work is wholeheartedly dedicated to my devoted family, whose unwavering encouragement, compassion, and understanding served as the impetus for finishing this senior project. Your unceasing support, selflessness, and faith in my abilities have kept me inspired throughout the difficult times. This accomplishment is a result of our team's efforts and the tightness of our bond. Thank you for always being there for me, for creating a safe space for me, and for teaching me the importance of perseverance and hard work.



ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to my family, whose unwavering support and encouragement have been the driving force behind the successful completion of my final year project. Their constant belief in my abilities, patience during my busy schedule, and understanding of the demands of this project have been invaluable. I am truly blessed to have such a loving and supportive family by my side.

I would also like to extend my sincere appreciation to my project supervisor, Prof. Dr. Mohd Faizal Abdollah, for the guidance, expertise, and unwavering commitment throughout this project. His valuable insights, constructive feedback, and dedication to my academic growth have immensely contributed to the quality and outcomes of this project. I am grateful for his time, patience, and mentorship, which have played a vital role in shaping my research and development skills.

This final year project would not have been possible without the collective support, guidance, and encouragement from my family and project supervisor. I am truly grateful for their contributions and the impact they have had on my academic and personal growth.

ABSTRACT

Authentication systems have been widely used in daily life for access control purposes. Unfortunately, the use of usernames and passwords to authenticate users is prone to various password attacks. Today, two-factor authentication systems which utilize both password and biometrics authentication systems can be used to mitigate these vulnerabilities. One of the biometrics types which is mostly unique for every person is keystroke dynamics. The typing patterns are based on the timing information of a key pressed and two successive key pressed. The goal of the project is to build a system which is able to record the keystroke patterns of 10 genuine users to determine if the users can be authenticated accurately by using machine learning. The machine learning algorithm used to build the system is Random Forest. The effectiveness of random forest in user authentication is tested and evaluated.

اونيورسيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

ABSTRAK

Sistem pengesahan telah digunakan secara meluas dalam kehidupan seharian untuk tujuan kawalan capaian. Malangnya, penggunaan pengguna dan kata laluan untuk mengesahkan pengguna terdedah kepada pelbagai serangan kata laluan. Pada masa kini, sistem pengesahan dua faktor yang menggunakan kata laluan dan sistem pengesahan biometrik boleh digunakan untuk mengurangkan kelemahan ini. Salah satu jenis biometrik yang kebanyakannya unik untuk setiap orang ialah dinamik ketukan kekunci. Corak menaip adalah berdasarkan maklumat masa ketukan ditekan dan dua ketukan berturut-turut ditekan. Matlamat projek adalah untuk membina sistem yang mampu merekodkan corak ketukan kekunci 10 pengguna tulen untuk menentukan sama ada pengguna boleh disahkan dengan tepat dengan menggunakan pembelajaran mesin. Algoritma pembelajaran mesin yang digunakan untuk membina sistem ialah Random Forest. Keberkesanan Random forest dalam pengesahan pengguna diuji dan dinilai.

TABLE OF CONTENTS

	PAGE
DECLARATION.....	II
DEDICATION.....	III
ACKNOWLEDGEMENTS.....	IV
ABSTRACT.....	V
ABSTRAK.....	VI
TABLE OF CONTENTS.....	VII
LIST OF TABLES.....	XIII
LIST OF FIGURES.....	XV
LIST OF ABBREVIATIONS.....	XVIII
CHAPTER 1: INTRODUCTION.....	1
1.1 Project Background.....	1
1.2 Problem Statement.....	2
1.3 Project Question.....	3
1.4 Project Objectives.....	3
1.5 Project Scope.....	3
1.6 Project Contribution.....	4
1.7 Report Organization.....	4

1.8	Conclusion	5
CHAPTER 2: LITERATURE REVIEW.....		6
2.1	Introduction.....	6
2.2	Authentication.....	6
2.3	Method of authentication	8
2.3.1	Text-based Authentication.....	8
2.3.2	Token-based Authentication.....	9
2.3.3	Biometrics-based Authentication.....	9
2.4	Keystroke Dynamics Authentication	10
2.4.1	Performance Parameter.....	11
2.4.2	Keystroke Dynamics Capture	14
2.4.2.1	Acquisition Hardware Device.....	15
2.4.2.2	Keystroke Dynamics Features	16
2.4.3	Benchmark Dataset for Keystroke Dynamics.....	19
2.4.4	Keystroke Dynamics Classification.....	19
2.4.4.1	Distance-based Approach	19
2.4.4.2	Machine Learning-based Approach.....	20
2.5	Critical review of existing algorithm and techniques	21
2.6	Proposed solution.....	26
2.6.1	Passphrase and password.....	28
2.6.1.1	Passphrase Condition.....	28
2.6.2	Supervised Machine Learning Algorithm	28
2.6.2.1	Random Forest.....	29
2.6.2.2	Support Vector Machine (SVM)	30

2.6.2.3	Multilayer Perceptron (MLP)	31
2.6.3	One-class Classification.....	33
2.6.3.1	Isolation Forest	33
2.6.4	Conceptual Architecture	34
2.7	Conclusion	36
CHAPTER 3: PROJECT METHODOLOGY		38
3.1	Introduction.....	38
3.2	Dataset.....	38
3.3	Project Methodology.....	39
3.3.1	Requirements Specification.....	40
3.3.1.1	Functional Requirement.....	41
3.3.1.2	Non-functional Requirement	42
3.3.1.3	Others Requirement	42
3.3.2	System Design	43
3.3.3	Detailed Design	43
3.3.4	Code and Test	43
3.3.5	Integration and Test	43
3.4	Project Milestones.....	44
3.5	Conclusion	46
CHAPTER 4: DESIGN		47
4.1	Introduction.....	47
4.2	Problem Analysis	47
4.3	Requirement Analysis.....	48

4.3.1	Data Requirement	48
4.3.2	Functional requirement	49
4.4	High-level Design	49
4.4.1	Data Collection	50
4.4.2	Data Preprocessing	52
4.4.3	Train Supervised Machine Learning Model	53
4.4.3.1	Random Forest Model	54
4.4.3.2	Support Vector Machines (SVM) Model	54
4.4.3.3	Multilayer Perceptron (MLP) Model	54
4.4.4	Classification	55
4.4.5	Evaluation	55
4.4.6	System Testing	55
4.5	Conclusion	57
CHAPTER 5: IMPLEMENTATION.....		58
5.1	Introduction.....	58
5.2	Implementation Status	58
5.3	Data Collection	60
5.3.1	User Interface.....	64
5.3.1.1	Registration.....	64
5.3.1.2	Login.....	65
5.3.2	Dataset	65
5.3.2.1	Training Dataset.....	65
5.3.2.2	Testing Dataset 1	66
5.3.2.3	Testing Dataset 2	66

5.4	Data Preprocessing.....	67
5.5	Machine learning model training, Classification and Evaluation	68
5.5.1	Training, Classification and Evaluation Process	68
5.5.1.1	Random Forest.....	68
5.5.1.2	Support Vector Machines (SVM).....	69
5.5.1.3	Multilayer Perceptron (MLP)	70
5.5.2	Machine Learning Model Training Result	71
5.6	System Testing.....	72
5.6.1	Anomaly Detection	72
5.6.2	Classifier.....	74
5.6.3	Integration of Isolation Forest and Random Forest	75
5.7	Conclusion	76
CHAPTER 6: TESTING		78
6.1	Introduction.....	78
6.2	Test Plan.....	78
6.2.1	Test Organization.....	78
6.2.2	Test Environment.....	79
6.2.3	Test Schedule.....	80
6.3	Test strategy	80
6.3.1	Classes of tests	81
6.4	Test Design	81
6.4.1	Performance Evaluation of Isolation Forest	82
6.4.2	Performance Evaluation of Random Forest.....	82
6.4.3	Usability Test of Keystroke Collection Process	83

6.4.4	Usability Test of Login Process.....	84
6.4.5	SQL Injection Test on Login Page	85
6.5	Test Result and Analysis.....	86
6.5.1	Performance Evaluation of Isolation Forest	86
6.5.2	Performance Evaluation of Random Forest.....	87
6.5.3	Usability Test of Keystroke Collection Process	88
6.5.4	Usability Test of Login Process.....	89
6.5.5	SQL Injection Test on Login Page	93
6.5.6	Overall Performance Evaluation of Machine Learning Algorithm	94
6.6	Conclusion	97
CHAPTER 7: PROJECT CONCLUSION		98
7.1	Introduction.....	98
7.2	Project Summarization.....	98
7.3	Project Contribution.....	99
7.4	Project Limitation	99
7.5	Future Works	100
7.6	Conclusion	100
REFERENCES.....		101

LIST OF TABLES

	PAGE
Table 1.1: Summary of Problem Statement	2
Table 1.2: Summary of Project Question.....	3
Table 1.3: Summary of Project Objectives	3
Table 1.4: Summary of Project Contribution.....	4
Table 2.1: Confusion Matrix	12
Table 2.2: Comparing Naive Bayes, Decision Trees and SVM on AUC and Accuracy (2005).....	14
Table 2.3: Final Results of Study based on Accuracy (2019).....	22
Table 2.4: Best Performance of Previous Studies	25
Table 2.5: Comparison of Techniques Used in Previous Studies	26
Table 3.1: Project Milestones PSM 1.....	44
Table 3.2: Project Milestones PSM 2.....	44
Table 3.3: Gantt Chart of PSM 1.....	45
Table 3.4: Gantt Chart of PSM 2.....	46
Table 4.1: Dataset description that depicts the information of the collected data.	48
Table 5.1: Implementation Status.....	58
Table 5.2: Summary of Machine Learning Model Training Result.....	72
Table 6.1: Test Organization.....	79
Table 6.2: Test Environment.....	79
Table 6.3: Test Schedule	80
Table 6.4: Classes of tests	81
Table 6.5: Test Design of Performance Evaluation of Isolation Forest Model... 	82
Table 6.6: Test Design of Performance Evaluation of Random Forest Model... 	83

Table 6.7: Test Design of Keystroke Collection Process.....	83
Table 6.8: Test Design of Login Process	84
Table 6.9: Test Design of SQL Injection Test on Login Page	86
Table 6.10: Summary of Performance Evaluation of Isolation Forest	87
Table 6.11: Summary of Performance Evaluation of Random Forest.....	88
Table 6.12: Summary of Usability Test of Keystroke Collection Process.....	89
Table 6.13: Summary of Usability Test of Login Process	92
Table 6.14: Summary of SQL Injection Test on Login page	94
Table 6.15: Overall Performance Evaluation of Machine Learning Algorithm	95



LIST OF FIGURES

	PAGE
Figure 2.1: Biometric-based Authentication	10
Figure 2.2: Keystroke Dynamics Authentication Approaches (Krishnamoorthy et al., 2018)	11
Figure 2.3: ROC curve (Fan et al., 2006)	13
Figure 2.4 Keystroke Dynamics Acquisition Hardware Device.....	16
Figure 2.5 : Keystroke Dynamics Features on Physical Keyboard	18
Figure 2.6 : Conceptual framework of random forest classifier (Parmar et al., 2019)	29
Figure 2.7: Kernel function helps SVM to deal with non-separable data (Noble, 2006)	31
Figure 2.8: Architecture of a Multilayer Perceptron (Gardner and Dorling, 1998)	32
Figure 2.9 : Conceptual Framework of Keystroke Dynamics Authentication System	36
Figure 3.1: Incremental Model (Graham, 1989)	40
Figure 4.1: Context Diagram for Keystroke Dynamics Authentication System	49
Figure 4.2: Flowchart of keystroke dynamics authentication system	50
Figure 4.3: Flowchart of Data Collection Process.....	51
Figure 4.4: Flowchart of Data Preprocessing	52
Figure 4.5: Flowchart of System Testing	56
Figure 5.1: Code snippet of calculation of dwell time and flight time.....	61
Figure 5.2: Code snippet of keyboard listener	62
Figure 5.3: Code snippet of passphrase validation	62
Figure 5.4: Code snippet of data export to MySQL.....	63

Figure 5.5: Code Snippet of User Interface for Registration	64
Figure 5.6: Code Snippet of User Interface for Login	65
Figure 5.7: Training Dataset, ‘keystroke’	66
Figure 5.8: Testing Dataset 1, ‘testset’	66
Figure 5.9: Testing Dataset 2, ‘mixset’	67
Figure 5.10: Code snippet of DataPreprocess.py module	68
Figure 5.11: Code snippet of Random Forest model during the training process	69
Figure 5.12: Random Forest model’s performance during training	69
Figure 5.13: Code snippet of SVM model during the training process	70
Figure 5.14: SVM model’s performance during training	70
Figure 5.15: Code snippet of MLP model during the training process	71
Figure 5.16: MLP model’s performance during training	71
Figure 5.17: Code snippet of load dataset ‘keystroke’ and ‘mixset’	73
Figure 5.18: Code snippet of preprocess test data from dataset ‘mixset’	73
Figure 5.19: Code snippet of Isolation Forest training and prediction	73
Figure 5.20: Export Random Forest Classifier	74
Figure 5.21: Code snippet of load Random Forest classifier and load dataset ‘keystroke’ and ‘testset’	74
Figure 5.22: Code snippet of preprocess test data from dataset ‘testset’	75
Figure 5.23: Code snippet of Random Forest classifier prediction	75
Figure 5.24: Code snippet of anomaly detection and classifier implementation	76
Figure 6.1: Test Result of Performance Evaluation of Isolation Forest	87
Figure 6.2: Test Result of Performance Evaluation of Random Forest	87
Figure 6.3: Keystroke Collection Interface for machine learning model training purposes	88
Figure 6.4: Keystroke Collection Interface for system testing purposes	89
Figure 6.5: Data collected	89
Figure 6.6: Login Interface	90
Figure 6.7: Successfully Logged In message	90
Figure 6.8: Error Message for staff and unknown user login attempt	91
Figure 6.9: Login attempts with invalid login credentials	92
Figure 6.10: Initial Login Attempt	93
Figure 6.11: SQL Injection Attempt	94

Figure 6.12: ROC curve of Random Forest..... 96

Figure 6.13: ROC Curve of Isolation Forest 97



LIST OF ABBREVIATIONS

FAR	-	False Acceptance Rate
FRR	-	False Rejection Rate
ROC	-	Receiver Operating Characteristic
EER	-	Equal Error Rate
RF	-	Random Forest
NN	-	Neural Network
SVM	-	Support Vector Machines
LR	-	Logistic Regression
DT	-	Decision Tree
MLP	-	Multilayer Perceptron
OCC	-	One-class Classification
IF	-	Isolation Forest
OCSVM	-	One-class Support Vector Machines
LOF	-	Local Outlier Factor

CHAPTER 1: INTRODUCTION

1.1 Project Background

User authentication system refers to a security mechanism employed to verify the identity of a user before granting the user's access to protected assets. There are multitude of methods to authenticate a user while the most common way is static method which involves the use of password or PIN. Although trustworthy, protected assets are vulnerable to those who gained access once a password or PIN is compromised. Some common password attacks which cause passwords to be compromised are phishing and social engineering, dictionary attacks, brute force, shoulder surfing, keylogging as well as database attacks. Even if a password is not compromised, a password authentication solution still provides a less efficient authentication. According to common password policies, employees should generate a strong password which is a long password with high entropy. Choong et al. (2014) state that employees spend a total of 18.6 hours or 2.25 days within 60 days on generating passwords for their work. Employees spend a lot of time in password generation and thus the time to manage these passwords can grow significantly. The strictness of password policy as security protocol facilitates the chance of login failures.

The use of dynamic authentication system is able to counter the issues. Keystore dynamics authentication is classified as behavioral biometric authentication method which is a part of dynamics authentication. Keystore dynamics authentication provides a backend solution that records the keystroke patterns of the users when they insert their text-based authentication into the system. Keystroke pattern recognition is a more efficient and more secure way to acknowledge the user since keystroke patterns

cannot be easily imitated. To overcome the problem of password authentication which subject to human memory error, passphrase is suggested to be added to the password policy as an additional option for users (Bhana and Flowerday, 2020). A user authentication solution using passphrases and keystroke dynamics can increase both security and usability.

A static keystroke dynamics authentication system specifically runs on an interface instead of continuously recording all the user's interactions with the system. Throughout this project, a static keystroke dynamics authentication system is designed and launched on a login screen. Passphrase is used as a replacement to complex password during authentication.

1.2 Problem Statement

Implementation of static user authentication using username and password to verify user is vulnerable to password attacks. Once a password is compromised, an attacker is authenticated by using the corresponding password and authorized to access protected assets. Besides, due to enforcement of password management and complexity policy, users are encouraged to use strong and complex passwords to mitigate the risk of password attack. A complex password might be difficult to memorize. The authentication process is subject to human memory error. Hence, a static keystroke dynamics authentication system which uses passphrase as login credentials is able to counter the issue. Keystroke dynamics authentication system validates user by checking the user's keystroke dynamics as well as the passphrase at the same time instead of checking at the password only. Keystroke dynamics is unique and not easy to copy. Thus, keystroke dynamics authentication system overcomes the limitation of password authentication system. A summary of problems that influence the motive of the project is shown in Table 1.1.

Table 1.1: Summary of Problem Statement

PS	Problem Statement
PS ₁	Static user authentication using username and password as login credentials is vulnerable to password attack.

1.3 Project Question

To counter the problems stated in the problem statement, there is a question that arise and need to be answered in this project. A summary of project questions is shown in Table 1.2.

Table 1.2: Summary of Project Question

PS	PQ	Project Question
PS ₁	PQ ₁	How keystroke dynamics authentication system helps in avoiding password attack?

1.4 Project Objectives

According to the problem statement and project question highlighted in Table 1.1 and Table 1.2, there are objectives that are aimed to be achieved in this project. These objectives are summarized in Table 1.3.

Table 1.3: Summary of Project Objectives

PS	PQ	PO	Project Objectives
PS ₁	PQ ₁	PO ₁	To investigate the use of static keystroke dynamics authentication as an access control mechanism.
		PO ₂	To develop a static keystroke dynamics authentication system using machine learning
		PO ₃	To test and validate the accuracy of static keystroke dynamics authentication system.

1.5 Project Scope

The scope of this project is listed as below:

1. To focus on the development and implementation of a static keystroke dynamics authentication system.

2. Normal user is the target user of this project
3. Passphrase from target user as login credential
4. Keystroke pattern of target user

1.6 Project Contribution

The expected outcome of this project which may contribute to the community is stated in Table 1.4.

Table 1.4: Summary of Project Contribution

PS	PQ	PO	PC	Project Contribution
PS ₁	PQ ₁	PO ₁	PC ₁	Proposed a suitable machine learning algorithm for keystroke pattern recognition.
		PO ₂		
		PO ₃	PC ₂	Proposed a static keystroke dynamics authentication system.

1.7 Report Organization

Chapter 1: Introduction

This chapter provides an overview of the project. It discusses the project background, problem statement, project question, objectives of the project, project scope, project contribution and report organization.

Chapter 2: Literature Review

This chapter shows how this project relates to existing research. It is mainly about a survey on scholarly sources related to the specific topic of this project. Strengths and weaknesses of past research on the specific topic are discussed and justification is made.

Chapter 3: Project Methodology