Universiti Teknologi Mara

Mobile Banking Trojan Detection Using Naïve Bayes

Anis Athirah Binti Masmuhallim

Thesis submitted in fulfilment of the requirements for Bachelor of Computer Science (Hons.) College of Computing, Informatics, And Mathematic

JANUARY 2024

ACKNOWLEDGEMENT

I am immensely grateful for the successful completion of this study within the allocated timeframe, and I attribute my achievement to the Grace and Blessing of Allah S.W.T. I extend my heartfelt gratitude to my advisor, Dr. Najiahtul Syafiqah Binti Ismail, whose unwavering support and valuable guidance made this research possible. Her dedication and contributions have been invaluable to the completion of this work.

I would also like to express my sincere appreciation to Dr. Ummu Fatihah Binti Mohd Bahrin, who served as my lecturer for CSP600 and CSP650. Her knowledge-sharing, encouragement, and inspiration throughout the semester played a crucial role in shaping this project. Undoubtedly, my family members have been my pillars of strength during this pursuit. I extend a special thank you to my beloved parents for their continuous support, both emotionally and financially, which was instrumental in overcoming challenges along the way.

Moreover, I am grateful for the collaborative efforts of all those involved in this project. My classmates deserve my deepest appreciation for their assistance and emotional support, which significantly contributed to the successful completion of my final-year project. Once again, I want to emphasize that this accomplishment would not have been possible without the help and encouragement of these remarkable individuals. Their support has been pivotal in my academic journey, and for that, I am truly thankful.

ABSTRACT

Mobile phones have become immensely popular as intelligent terminals worldwide. The open-source nature of mobile platforms has facilitated the development of third-party mobile applications, but it has also created an environment for mobile malware to thrive. Unfortunately, the abundance of mobile applications and lax management of some app stores has led to potential risks for mobile users, including privacy breaches and malicious deductions of fees, among other adverse consequences. This research presents a mobile malware static detection method based on Naïve Bayes. The approach aims to offer a solution to protect users from potential threats such as Banking Trojan malware. The objectives of this project are to study the requirement of the Naïve Bayes algorithm in Mobile Banking Trojan detection, to develop a web-based detection system for Mobile Banking Trojan using Naïve Bayes, and to evaluate the performance and accuracy of the Naïve Bayes algorithm in the Mobile Banking Trojan detection. The methodology of this project consists of the research framework, preliminary study, design, development, and performance evaluation.

TABLE OF CONTENTS

CONTENT **PAGE**

SUPERVISOR APPROVAL ii

STUDENT DECLARATION iii

ACKNOWLEDGEMENT iv

ABSTRACT v

TABLE OF CONTENTS vi

LIST OF FIGURES x

LIST OF TABLES xii

LIST OF ABBREVIATIONS xiii

CHAPTER 1: INTRODUCTION 1

1.1 Background of Study 1

1.2 Problem Statement 3

1.3 Objectives 5

1.4 Project Scope 6

1.5 Project Significance 7

1.6 Overview of Research Framework 8

1.7 Conclusion 10

CHAPTER 2: LITERATURE REVIEW 11

2.1 Mobile Malware 11

2.2 Mobile Malware Detection 13

2.3 Mobile Banking Trojan Detection 14

2.3.1 Banking Trojan 14
2.3.2 Characteristics of Banking Trojan 15
2.3.3 How the Banking Trojan Affects Mobile User 16
2.3.4 Benefits of Early Detection of Banking Trojan 18
2.4 Naïve Bayes 19
2.4.1 Overview of Naïve Bayes 19
2.4.2 Naïve Bayes and How It Works 19
2.4.3 Naïve Bayes Algorithm Feature & Characteristic 23
2.4.4 Gaussian Naïve Bayes Classifier 23
2.4.5 Advantages of Naïve Bayes Algorithm 24
2.4.6 Implementation of Naïve Bayes in Various Problems 26
2.6 Similar Work 32
2.7 Implication of Literature Review 37
2.8 Chapter Summary 38

CHAPTER 3: METHODOLOGY 39
3.1 Overview of Research Framework Methodology 39
3.1.1 Detailed Research Framework 39
3.2 Preliminary Study 42
3.2.1 Literature Study 42
3.2.2 Data Collection 43
3.2.3 Data Pre-processing 45
3.3 Design 47
3.3.1 Software Development Life Cycle (SDLC) 48
3.3.2 System Architecture 49
3.3.3 Flowchart 50
3.3.4 Pseudocode 51