SIM BOX DETECTION USING MACHINE LEARNING TECHNIQUES

Weerakkodi Pathirathnage Tharindu Darshana Dissanayake

(188252k)

Degree of Master of Science

Department of Electronic and Telecommunication Engineering

University of Moratuwa

Sri Lanka

January 2022

SIM BOX DETECTION USING MACHINE LEARNING TECHNIQUES

Weerakkodi Pathirathnage Tharindu Darshana Dissanayake

(188252k)

Dissertation submitted in partial fulfillment of the requirements for the degree
Master of Science

Department of Electronic and Telecommunication Engineering

University of Moratuwa
Sri Lanka

January 2022

DECLARATION

Candidate:

I declare that this is my work to the best of my knowledge, and this report does not incorporate any material previously submitted for a degree or diploma in any university or institute of higher learning without acknowledgment. This report does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

I also hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my report, in whole or in part in print, electronic, or any other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

……………….…………… ……………………….

Dissanayake W. P. T. D Date

Supervisor:

I certify that the declaration above by the candidate is true to the best of my knowledge, and the candidate has carried out research under my supervision.

……………….…………… ..…………………….

Dr. Ranga Rodrigo Date

Abstract

Subscriber Identity Module (SIM) box fraud causes severe International Direct Dialing (IDD) revenue losses to network operators and also causes harm to their brand names due to its low Quality of Service (QoS). Therefore, it is very important to prevent this fraud by disconnecting SIM numbers used for SIM box fraud with smaller call attempts to hinder fraudsters from generating revenue. However, detection of SIM box fraud numbers for disconnection consumes lots of work hours as it is necessary to minimize the impact on genuine subscribers. Call Detail Record (CDR) analysis is one of the methods to detect SIM numbers used in the SIM box. Machine learning techniques for fast data analysis are suitable for detecting SIM box numbers using CDR. This research was carried out to accurately identify suitable machine learning methods to detect SIM box fraud numbers. Subscriber profiles are created using CDR to train and evaluate machine learning models. Support vector classifier, decision tree classifier, and K-nearest neighbors classifier algorithms show better results from different types of machine learning models that were trained and evaluated to identify suitable algorithms to detect accurately. A weighted linear combination of these algorithms was used to get the best result with a macro F1-score value of 0.89 using the ensemble voting classifier algorithm. The result confirms that machine learning models can be used to detect SIM box fraud numbers accurately.

Keywords: International direct dialing, SIM box fraud, bypass fraud, machine learning, neural networks, support vector machine algorithms, decision tree algorithms, K-nearest neighbors algorithms, ensemble voting algorithm.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to Dr. Ranga Rodrigo for the guidance, supervision, and advice given to complete this research. Then, I would also like to extend my gratitude to Dr. Tharaka Samarasinghe, Dr. Kasun Hemachandra, and other lecturers who provide guidance and support to complete the M.Sc. in Telecommunication and the research. Next, I would like to thank my supervisors from Dialog Axiata PLC Mr. Dasun Kehelwala, and Mr. Sampath Ilesinghe for providing help and approval to complete the research successfully. Finally, I would like to thank my family members and friends who helped me a lot in completing the research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CDR	Call Detail Record
CFCA	Communications Fraud Control Association
CLI	Calling Line Identification
IDD	International Direct Dialing
IMEI	International Mobile Equipment Identity
IMSI	International Mobile Subscriber Identity
IP	Internet Protocol
IRSF	International Revenue Share Fraud
MO	Mobile Originated
MT	Mobile Terminated
PBX	Public Branch Exchange
QoS	Quality of Service
SIM	Subscriber Identity Module
SIP	Session Initiation Protocol
SMS	Short Message Service
USD	United States Dollars
VoIP	Voice over Internet Protocol