DEVELOPMENT OF A FEDERATED LEARNING-BASED MALWARE DETECTION MODEL FOR INTERCONNECTED CLOUD INFRASTRUCTURES

MUGHOLE, KALIMUMBALO DANIELLA
20PCK02091

MARCH, 2023

DEVELOPMENT OF A FEDERATED LEARNING-BASED MALWARE DETECTION MODEL FOR INTERCONNECTED CLOUD INFRASTRUCTURES

BY

MUGHOLE, KALIMUMBALO DANIELLA
(20PCK02091)
B.Eng. Génie Electrique et Informatique,
Université Libre des Pays des Grands Lacs, Goma, DR Congo

A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF ENGINEERING DEGREE (M.ENG.) IN INFORMATION AND COMMUNICATION ENGINEERING, DEPARTMENT OF ELECTRICAL AND INFORMATION ENGINEERING, COVENANT UNIVERSITY, OTA, OGUN STATE, NIGERIA

MARCH, 2023

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfilment of the requirements for the award of Master of Engineering degree (M.Eng) in Information and Communication Engineering, Department of Electrical and Information Engineering, College of Engineering, Covenant University, Ota, Ogun State, Nigeria.

Miss Adefunke F. Oyinloye
(Secretary, School of Postgraduate Studies) **Signature and Date**

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies) **Signature and Date**

DECLARATION

I, **MUGHOLE, KALIMUMBALO DANIELLA (20PCK02091)**, declare that this research was carried out by me under the supervision of Dr. Joke A. Badejo and Dr. Kennedy O. Okokpujie of the Department of Electrical and Information Engineering, College of Engineering, Covenant University, Ota, Ogun State, Nigeria. I attest that the dissertation has not been presented either wholly or partially for the award of any degree elsewhere. All sources of data and scholarly information used in this dissertation are duly acknowledged.

MUGHOLE, KALIMUMBALO DANIELLA

Signature and Date

CERTIFICATION

We certify that this dissertation titled " **DEVELOPMENT OF A FEDERATED LEARNING-BASED MALWARE DETECTION MODEL FOR INTERCONNECTED CLOUD INFRASTRUCTURES**" is an original research work carried out by **MUGHOLE, KALIMUMBALO DANIELLA (20PCK02091)**, in the Department of Electrical and Information Engineering, College of Engineering, Covenant University, Ota, Ogun State, Nigeria, under the supervision of Dr. Joke A. Badejo and Dr. Kennedy O. Okokpujie. We have examined and found this work acceptable as part of the requirements for the award of Master of Engineering in Information and Communication Engineering.

Dr. Joke A. Badejo
(Supervisor) **Signature and Date**

Dr. Kennedy O. Okokpujie
(Co-Supervisor) **Signature and Date**

Prof. Emmanuel Adetiba
(Head of Department) **Signature and Date**

Dr. Oluwumi Adetan
(External Examiner) **Signature and Date**

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies) **Signature and Date**

DEDICATION

This dissertation is first and foremost dedicated to God Almighty, the source of all wisdom, knowledge, and understanding, for His grace and favour throughout this research. Then to my parents KAMBALE KALIMUMBALO and KABUO ISEGHUNDI for their endless support and love.

ACKNOWLEDGMENTS

My profound gratitude goes firstly to God Almighty for granting me the willpower to pursue this study and for imparting the wisdom, strength, support, knowledge, and health to do so, most pertinent of which is life itself.

The expression of great gratitude is addressed to my supervisor Dr. Joke A. Badejo, for her support, comments, and encouragement. Also, my deep thanks to the management of Covenant Applied Informatics and Communications - African Centre of Excellence (CApIC-ACE) for the incredible support throughout my master's studies, particularly to the director Professor Ezekiel Adebiyi and the deputy director Prof. Emmanuel Adetiba.

My deep thanks also go to Prof. Emmanuel Adetiba, Dr. Kennedy Okokpujie, Dr. Victoria Oguntosin, Prof. Francis Idachaba, and Dr. Oluwadamilola Oshin, for support, corrections, and above all encouragement. I especially appreciate all my lecturers for all the labor put into teaching and mentoring me and my colleagues through all the courses they handled. I also appreciate the effort put into developing and perfecting this research through constructive criticism and feedback.

My appreciation goes as well to Prof. Takenga Claude for his support, encouragement, and advice throughout my studies.

My sincere thanks remain meritorious to my parents, sisters, brothers, and friends who helped me achieve my goals; I quote Fleurette Vivuya, Raph-Arsène Kasongo, Inès Muheruki, Justine Kahindo, Emile Kalimumbalo, Josiane Ndaghane, and Elie Lubamba. Thank you for supporting me throughout my studies and especially praying and encouraging me on all those very hard days.

May my friends Molo Mbasa, Patrick Vingi, Myriam Baiguerel, John Chishugi, Sonia Nkongho, Mariam Kanonte, and Lumbani Bondera, find here the expression of my gratitude for their support, friendship, jokes, encouragement, and prayers.

TABLE OF CONTENTS

CONTENTS **PAGES**

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Networks
CApIC-ACE	Covenant Applied Informatics and Communication Africa Centre of Excellence
CC	Cloud Computing
CCMP	Cloud Computing Management Platforms
CSC	Cloud Service Consumer
CSP	Cloud Service Provider
CSU	Cloud Service User
CSV	Comma Separated Value
DL	Deep Learning
FedAvg	Federated Averaging
FEDGEN	Federated Genomic Cloud
FFNN	Feedforward Neural Network
FL	Federated Learning
FLOWER	Friendly Federated Learning Research Framework
FLS	Federated Learning System
FN	False Negative
FP	False Positive
HFL	Horizontal Federated Learning
IaaS	Infrastructure as a Service
IDS	Intrusion Detection System
LSTM	Long Short-Term Memory
MDS	Malware Detection System
ML	Machine Learning

MLP	Multi-Layer Perceptron
PaaS	Platform as a Service
PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
SaaS	Software as a Service
SMOTE	Synthetic Minority Oversampling Technique
TN	True Negative
TP	True Positive
VFL	Vertical Federated Learning
VM	Virtual Machine

ABSTRACT

Due to the large number of heterogeneous applications using the same infrastructure, enforcing security and reliability in the cloud is a difficult but crucial task. A security analysis system that detects threats for example malicious software (malware) should exist within the cloud infrastructure. Different malware techniques that bypass network-based and host-based security protections have led to the development of new methods for analysing and detecting malware, which have evolved over the past decades. Due to the complexity of learning the ever-changing configurations of malware, it is challenging for forensics investigators to keep up with the exponential rise in the number and variety of malware species. In this research work, a malware detection model was developed for interconnected cloud infrastructures based on federated learning. This technique enables collaboration between multiple devices in the training of machine learning models without exchanging data, thereby preserving the privacy of individual users. Three different deep-learning algorithms were selected and used in the training, validation, and testing of the models. By the model training with eight clients and twenty-five federation rounds, the FeedForward Neural Networks(FFNN) model provided the best performance with a precision of 84%, an F1-score of 84%, and an accuracy of 84% whereas the Multi-Layer Perceptron(MLP) model provided 83% of precision, 83% of F1-score, and 83% of accuracy and the Long Short-Term Memory(LSTM) model provided a performance with 80% of precision, 80% of F1-score, and 80% of accuracy as well.

Keywords: Federated learning, malware detection, federated cloud, machine learning