AN EFFICIENT DETECTION MODEL OF ZERO-DAY WEB APPLICATION ATTACKS BASED ON CONVOLUTION NEURAL NETWORKS AND DEEP AUTOENCODERS

TUEI KEVIN KIRUI

A Thesis Submitted to the Graduate School in Partial Fulfillment of the Requirements for the Award of the Degree of Master of Science in Computer Science of Chuka University.

CHUKA UNIVERSITY

OCTOBER, 2024

DECLARATION AND RECOMMENDATION

Declaration

This thesis is my original work and has not been submitted for examination in any other University

Signature_____ Date 14/10/2024

Kevin Kirui Tuei
SM22/39984/19

Recommendation

This thesis has been examined, passed and submitted with our approval as University supervisors

Signature_____ Date_____

Dr. David G. Mwathi, PhD
Chuka University

Signature_____ Date 14/10/2024

Dr. Edna C. Too, PhD
Chuka University.

DEDICATION

There is a success, then there is the Good Success. This long and winding journey would not be possible without the support of my family, my siblings Nancy, Cynthia and most importantly my father Stephen Tuei, and my mother Catherine Tuei who have been my tower of strength and source of hope. Special dedication goes to my colleague Rose with whom we walked this postgraduate journey. I dedicate this effort to all of you.

ACKNOWLEDGMENT

First of all am grateful to the Almighty God for the grace to commit to the completion of this thesis. I would like to express my sincere gratitude to my supervisors Dr. David Mwathi and Dr. Edna Chebet for their relentless support. The assistance they provided in my research endeavour was phenomenal.

I also take this opportunity to acknowledge the support of the Academic Staff in the Faculty of Science, Engineering and Technology who took their time to review my work and support me through the process of both proposal and thesis writing.

This research was also made possible through the contributions of many other supportive personalities all of whom I may not be able to name and I really acknowledge them for their contribution towards this journey.

Special recognition to my amazing partner and friends.

ABSTRACT

The need for secure and trustworthy information systems has taken center stage and proven critical in supporting teleworking, online teaching, and research services. Artificial Intelligence (AI) is the primary driver of the 6th generation of computing, and innovations with applications of AI in computer vision, gaming, robotics, and security. Zero-day web application attacks take advantage of web application software weakness for as long as the developer is unaware and has not developed a mechanism to eliminate the weakness. Zero-day attacks leave vulnerable users grappling with data loss and have the propensity to push an organization out of business. Current zero-day attack detection methods built on signature-based or anomaly-based methods are inefficient in combating these attacks since they rely on previously detected weaknesses for signatures and a deviation from normal behavior for anomaly detection. These methods result in detection rates below 80%, meaning the propensity of Zero-day attacks going undetected is 20% or lower. The application of machine learning techniques has proven to be efficient because these techniques can continuously learn from the code as well as its execution to detect security breaches and trigger an alarm. With the need to improve these techniques, a novel classification model needs to be developed to increase the detection rate further and reduce the false alarm rate. This study applied a hybrid of two machine learning methods, Convolution Neural Networks and deep autoencoders, to develop a classification model that significantly increases the detection rate of zero-day attacks. The KDD'99 Dataset is a comprehensive repository of fully labeled intrusion detection records that was used to develop, test and validate the model. This dataset simulated real-world scenarios and assessed the model's performance under different intrusion scenarios. The Average Detection Rate, Accuracy and F1 score metrics were used to evaluate the model. The hybrid CNN-Deep Autoencoder model had a detection rate of 0.895 against 0.887 of the Fully Connected Network (FCN) with sampling and 0.885 of the pure CNN model. The accuracy and
F1-score of the hybrid CNN-Deep Autoencoder were 0.973 and 0.971 respectively. The
Hybrid Model of CNN and Deep Autoencoder is efficient in detecting Zero-Day Attacks making it possible for Software Developers to patch their systems sooner resulting in minimal dwell time.