A Summary of Thesis
submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy
IN
COMPUTER SCIENCE

Submitted by

Arvind Prasad

Enrolment No.: 1529/19

Under Supervision of

Dr. Shalini Chandra

Submitted to

DEPARTMENT OF COMPUTER SCIENCE
SCHOOL FOR INFORMATION SCIENCE AND TECHNOLOGY
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY)
VIDYA VIHAR, RAE BARELI ROAD
LUCKNOW-226025, UTTAR PRADESH, INDIA

2023

ABSTRACT

The widespread adoption of the Internet and Internet-connected devices has revolutionized the way we engage with daily services such as education, finance, healthcare, and shopping through the Internet. This digital transformation has significantly improved the convenience and efficiency of our lives. However, our growing reliance on devices like laptops, smartphones, and smart home technology has made our personal information increasingly vulnerable. As technology advances and our dependence on digital systems deepens, the need to secure our sensitive information and protect devices in the digital world becomes essential. In this digital age, enhancing security against cyber threats is highly important.

This research discuses and highlights the imperative need to tackle the escalating threat posed by multi-vector cyberattacks. The prevalence of cyber threats, including Distributed Denial of Service (DDoS), phishing, botnet, and malware attacks, is discussed in depth. These threats pose a diverse range of challenges, from disrupting online services and stealing sensitive information to undermining user trust. As the digital landscape continues to evolve, the need for advanced defense mechanisms becomes increasingly evident.

Machine learning, a subfield of artificial intelligence, emerges as a promising solution in the quest for cybersecurity enhancement. The researchers outline the stages of machine learning, including data preprocessing, feature selection, hyperparameter tuning, model selection, training, and evaluation. It provides a structured framework for understanding the application of machine learning in cybersecurity. The chapter also explores various machine learning techniques, such as ensemble techniques (bagging and boosting),

stacking ensemble techniques, and voting techniques, as well as incremental learning techniques. These techniques offer the potential to combat evolving cyberattacks and enhance cybersecurity measures.

Furthermore, the researcher conducted an exhaustive literature review, laying the groundwork for the foundation of this research. This review surveys the state-of-the-art endeavors in machine learning-based security techniques. These techniques are designed to safeguard against a wide spectrum of cyberattacks. The review highlights the incredible potential of machine learning, which has inspired numerous researchers to develop approaches for detecting and mitigating cyber threats. However, it also underscores the challenges and research gaps in the field, emphasizing the need for comprehensive and innovative techniques to address the evolving nature of cyberattacks.

Volumetric DDoS attacks is critical for online security as these attacks pose a significant threat to online service availability, potentially leading to financial losses, credibility damage, and user trust erosion. The researchers have developed VMFCVD, a framework to combat volumetric DDoS attacks. VMFCVD introduces a novel approach with three modes: Fast Detection Mode (FDM), Defensive Fast Detection Mode (DFDM), and High Accuracy Mode (HAM). FDM quickly classifies network traffic during attacks with reduced data dimensions. DFDM enhances malicious traffic detection, while HAM prioritizes accuracy. The framework also underscores the significance of data preprocessing and reduced feature set identification steps to enhance detection efficiency.

Defending against phishing attacks is crucial as they exploit human vulnerability to deceive individuals into disclosing sensitive information, posing a severe risk to data security, financial assets,

and personal privacy. The researchers introduce the PhiUSIIL framework, a phishing URL detection framework based on similarity index and incremental learning. This framework introduces a diverse dataset, including legitimate and phishing URLs, enhancing model generalization and reducing bias. It incorporates a URL similarity index technique to detect visually similar attacks, which can deceive users into accessing malicious sites. Incremental learning and diverse security profiles are integrated into the framework to improve scalability, adaptability, and model resilience in phishing URL detection.

Further, the researchers developed the BotDefender framework, a collaborative defense framework against botnet attacks. Botnet attacks have become a significant cybersecurity menace, posing threats to individuals, governments, and businesses. BotDefender introduces a two-stage botnet attack detection technique that combines network traffic analysis and machine learning capabilities. This approach efficiently identifies and discards a significant portion of botnet-related traffic through the network traffic analyzer, reducing the computational workload on the subsequent machine learning model. The extended dataset construction technique enhances the model's learning capability, and feature selection prioritizes important features to improve detection accuracy. Real-world botnet attack strategies are employed to evaluate the effectiveness of BotDefender.

The researchers have developed the AndroMD framework as an Android malware detection system. The AndroMD framework creates an extensive dataset of 600,298 unique Android apps that provides a benchmark dataset for research in the field of Android malware. The research introduces the AndroMD optimal feature selection (AOFS) technique, which efficiently selects and aggregates top features to

create an optimal feature subset. It explores aggregator-based malware detection using the AndroMD datasets and conducts live experiments to validate the approach under real-world conditions. To validate the malware detection approach, the study developed eight legitimate and malware apps and conducted live experiments. This practical testing underscores the real-world applicability and robustness of the proposed approach.

In conclusion, this thesis explores the dynamic intersection of cybersecurity and machine learning, offering innovative frameworks and techniques to defend against various cyberattacks. The research contributions presented by the researchers provide practical and effective solutions to enhance security in an ever-evolving digital landscape. Machine learning's adaptability and potential to learn from data and adapt to emerging threats make it a valuable tool in the ongoing battle to protect sensitive data and digital infrastructure. These frameworks represent not only significant advancements in cybersecurity but also valuable resources for future research and cybersecurity efforts.