Intelligent Anomaly Detection System Based on Ensemble and Deep Learning

Babu Kaji Baniya, Thomas Rush

Department of Computer Science & Information Systems, Bradley University, Peoria, IL, United States

bbaniya@fsmail.bradley.edu, trush@mail.bradley.edu

***Abstract*— The ubiquity of the Internet plays a pivotal role in connecting individuals and facilitating easy access to various essential services. As of 2022, the International Telecommunication Union (ITU) reports that approximately 5.3 billion people are connected to the internet, underscoring its widespread coverage and indispensability in our daily lives. This expansive coverage enables a myriad of services, including communication, e-banking, e-commerce, online social security access, medical reporting, education, entertainment, weather information, traffic monitoring, online surveys, and more. However, this open platform also exposes vulnerabilities to malicious users who actively seek to exploit weaknesses in the virtual domain, aiming to gain credentials, financial benefits, or reveal critical information through the use of malware. This constant threat poses a serious challenge in safeguarding sensitive information in cyberspace. To address this challenge, we propose the use of ensemble and deep neural network (DNN) based machine learning (ML) techniques to detect malicious intent packets before they can infiltrate or compromise systems and applications. Attackers employ various tactics to evade existing security systems, such as antivirus or intrusion detection systems (ADS), necessitating a robust defense mechanism. Our approach involves implementing an ensemble, a collection of diverse classifiers capable of capturing different attack patterns and better generalizing from highly relevant features, thus enhancing protection against a variety of attacks compared to a single classifier. Given the highly unbalanced dataset, the ensemble classifier effectively addresses this condition, and oversampling is also employed to minimize bias toward the majority class. To prevent overfitting, we utilize Random Forest (RF) and the dropout technique in the DNN. Furthermore, we introduce a DNN to assess its ability to recognize complex attack patterns and variations compared to the ensemble approach. Various metrics, such as classification accuracy, precision, recall, F1-score, confusion matrix are utilized to measure the performance of our proposed system, with the aim of outperforming current state-of-the-art intrusion detection systems..**

***Keyword*— cybersecurity, deep neural network, ensemble, generalizing**

Babu Kaji Baniya holds a B.E. degree in Computer Engineering from Pokhara University, Nepal, which he obtained in 2005. He further pursued his education and completed an M.E. and Ph.D. in Department of Computer Science and Engineering from Chonbuk National University, Republic of Korea in 2015. Following his doctoral studies, he gained valuable experience as a postdoctoral researcher in the Department of Computer Science and Biomedical Engineering at the University of South Dakota. He then served as an assistant professor in the Department of Computer Science and Digital Technologies at Grambling State University, Louisiana. Currently, he holds the position of assistant professor in the Department of Computer Science and Information Systems at Bradley University, located in Peoria, Illinois. Throughout his career, he has taught a wide range of Computer Science courses at both the graduate and undergraduate levels. His research interests span several key areas, including audio signal processing, information retrieval, cybersecurity, bioinformatics, Big Data, and machine learning. A specific focus of his research involves the application of machine learning and deep learning algorithms in securing the Internet of Medical Things (IoMT). He is also IEEE member.

Thomas Rush is currently pursuing a B.S. degree in Computer Science with a concentration in Data Science at Bradley University in Peoria, Illinois. His research experience includes working as a Research Assistant to Dr. Baniya at Bradley University, where he focused on classifying Android malware using a variety of machine learning models and deep learning techniques. Thomas has applied these skills in practical settings, having completed coursework in Data Science, Machine Learning, Data Mining, and Artificial Intelligence at Bradley University. Additionally, he contributed to data analytics for the Bradley University Men's Soccer Team through his senior capstone project.