Review of Machine Learning Applications in Enhancing Software Security

Students Name

University

Course

Professor

Date

Review of Machine Learning Applications in Enhancing Software Security

Numerous internet protocols are applied by programs that utilize different user capacities. The primary system of communication utilized by these protocols is data packets. Network data relayed and collected at both wireless and physical interfaces can be seized and accumulated in pcap (packet capture) format. Therefore, popular applications such as Wincap and Libpcap are applied in analyzing network traffic in Windows and UNIX, respectively.

In contrast to network data packets, machine learning datasets possess peculiar characteristics and functionalities. These special features establish the primary attributes of every data class in the data set. Consequently, whenever unrefined data is seized as a pcap, the researcher has to generate a particular script to isolate the properties required from the pcap and convert them into a machine learning applicable format. Alothman (2019) analyzed the arff (attribute relation file format) of Weka and formulated a mechanism that can be applied to transform pdml (Packet Details Markup Language) into a Weka arff configuration. In particular, the researcher utilized a tshark application to create a pdml file from a pcap format. He established that the tshark tool is effective in transforming unrefined traffic into a Weka-suitable variant.

Zor et al. (2018) acknowledge the utilization of real-time signature disclosure through the application of a clustering algorithm. The researchers applied two clustering systems using a density-dependent clustering system referred to as SLCT (Simple Logfile Clustering Tool) to detect both ordinary and invasion events as well as determine regular traffic. The researchers noted that the accuracy of invasion detection was elevated to approximately 82% when cluster integrity was utilized as a performance benchmark. Therefore, this machine learning clustering

technique attained an impressive degree of accuracy regardless of the zero-day nature of the invasions.

Picek et al. (2018) utilized an ordinary GP (Genetic Algorithm) system to create an invasion classification system. Genetic algorithm and Genetic programming are some of the most widely utilized computation techniques designed on the basic rule of survival of the fittest. Notably, these algorithms rely on the roles of chromosome categories that advance depending on particular operators. The algorithm is initiated using a randomly created group that applies a fitness value to benchmark the capacity of every individual in solving an existing problem. The researchers utilized three common GP systems to generate the classification scheme. They include the MEP (Multi Expression Programming), GEP (Gene Expression Programming), and LGP (Linear Genetic Programming). In particular, the researchers employed the Defense Advanced Research Project Agency (DARPA) data set to authenticate the produced model. Accordingly, this machine learning technique produced a decreased FAR (False Alarm rate) ranging from five to zero percent. Therefore, the application of the GP machine learning technique significantly decreases the False Alarm Rate in cybersecurity platforms.

Naseer et al. (2018) applied an ANN (Artificial Neural Network) system with various category classifications to disclose anomalies. The ANN technique works similarly to a human brain whereby input from the nerves triggers neurons in the secondary periphery of the brain networks. Correspondingly, the created output is relayed to the secondary hierarchy layer until the final output is generated by the last network zone. Therefore, the crucial network elements that play significant roles in neural connections are shielded from environmental elements. The primary disadvantage of the ANN system is its vast learning duration as a result of the existing residential minima. With this in mind, the researchers produced data using the RealSecure

network analyzer and simulated approximately three thousand invasions using software such as Internet Scanner. The researchers then pre-refined the acquired data using nine attributes such as raw data category and span, destination and source port, protocol recognizer, destination, and source location, ICMP category, and ICMP code. Notably, the researchers noted an error percentage of 0.07 and 0.058 during the evaluation and training phases, respectively. Therefore, the ANN technique of machine learning elevated the efficiency of anomaly disclosure by about 93%.

References

Alothman, B. (2019). Raw Network Traffic Data Preprocessing and Preparation for Automatic Analysis. *2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*. https://doi.org/10.1109/cybersecpods.2019.8885333

Naseer, S., Saleem, Y., Khalid, S., Bashir, M. K., Han, J., Iqbal, M. M., & Han, K. (2018). Enhanced Network Anomaly Detection Based on Deep Neural Networks. *IEEE Access*, *6*, 48231–48246. https://doi.org/10.1109/access.2018.2863036

Picek, S., Hemberg, E., Jakobovic, D., & O'Reilly, U.-M. (2018). One-Class Classification of Low Volume DoS Attacks with Genetic Programming. *Genetic Programming Theory and Practice XV*, 149–168. https://doi.org/10.1007/978-3-319-90512-9_10

Zor, C., Kittler, J., Wang, W., Kaloskampis, I., Hicks, Y., & Hunter, A. 5.3 Automated statistical anomaly detection and incongruence determination. *In Signal Processing*, 148.