UNIVERSITI TEKNOLOGI MARA

DDoS DETECTION AND DEFENSE MECHANISM BASED ON PACKET ANALYSIS AND MACHINE LEARNING

MOHD AZAHARI BIN MOHD YUSOF

Thesis submitted in fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Computer Science)

Faculty of Computer and Mathematical Sciences

November 2019

ABSTRACT

DDoS attacks are one of the most serious threats nowadays. It is an attack launched by an attacker using various source of IP addresses to disrupt a network or IoT environment. There are several types of DDoS attacks such as TCP SYN flood, UDP flood, ICMP/Ping flood, Ping of Death, Smurf, Slowloris, HTTP flood, Zero-day attack and SIDDoS which can be launched by attackers for a specific purpose. Therefore, this research is aimed at overcoming high false positive rate to improve detection accuracy against several DDoS attacks. This research focused on four types of DDoS attacks, namely TCP SYN flood, UDP flood, Ping of Death and Smurf. Packet Threshold Algorithm (PTA) was introduced in this research, where the technique can detect incoming packets either normal packets or DDoS attacks. The packet class detected by the PTA is based on the specified packet threshold. All types of DDoS attacks that have been detected by the PTA will be dropped and will be stored in the log of packets. Meanwhile, normal packets that have been detected by PTA are allowed into the network or IoT environment. Then, this technique is coupled with four machine learning, namely SVM, Naïve Bayes, Logistic Regression and KNN. Finally, they are evaluated based on detection accuracy, FP rate, TP rate, TN rate, FN rate, precision, recall, f1-score and total of running time to see the efficiency of feature selection methods for reducing false positive rate in DDoS attack reduction. Based on performance comparison, it shows that PTA-KNN technique has achieved 99.83% detection accuracy with only 0.02% FP rate, which is considered the best technique to reduce false positive rate problem.

ACKNOWLEDGEMENT

Firstly, I wish to thank God for giving me the opportunity to embark on my PhD and for completing this long and challenging journey successfully. My gratitude and thanks go to my supervisor Dr. Fakariah Hani Binti Mohd Ali and Dr. Mohamad Yusof Bin Darus for valuable guidance, constant support and intuitive supervision.

I also do not forget to thank the lecturers, officers and staff of Faculty of Computer and Mathematical Sciences (FSKM) and Institute of Graduate Studies (IGS) of Universiti Teknologi MARA (UiTM) Shah Alam for their assistance during this research.

I would also like to thank my fellow friends who had given encouragement and support to me. Finally, I would like to convey thanks to my parents, wife and siblings who have supported me in the process of completing this thesis.

TABLE OF CONTENTS