Feature Engineering for Distributed Denial of Service (DDoS) attack classification

Alan Barzilay and Caio L. Martinelli

April 2020

1 Introduction

DDoS (Distributed Denial of Service) attacks have always been a menace for service providers everywhere, but with the increase of IoT devices came also a number of new opportunities for attackers. The Mirai botnet(1) is only one example of such opportunities, where a botnet of vulnerable IoT devices was used to create a massive DDoS attack. With the possibility of new and powerful attacks being ever present, the need for efficient and agile tools and models for DDoS detection is a must if we ever intend to be able to take countermeasures. But the massive volume of data being generated and transmitted everyday imposes a serious challenge, most common and readily available tools aren't able to keep up with the sheer volume of information being received by a usual server. In order to develop new models and tools we need to be capable of manipulating the data stream and easily extract features that are descriptive of the flux being analyzed.

In this work we intend to develop a new set of features that are computationally inexpensive and descriptive of the data stream. We expect to generate this set by reviewing the literature for features already in use and develop new ones. Once we have this set we will compare its descriptive power with other sets of features used in other models.

For this we will use the realistic DDoS dataset described on the article (2), which gather a variety of attacks, including very up-to-date techniques. We hope to extract relevant features to classify the bidirectional flows, as the dataset article did, using the common definition of flow, a sequence of packets with same values for {*Source IP, Destination IP, Source Port, Destination Port and Protocol (TCP or UDP)*}, the same one used on it.

The feature generator used in it to evaluate the power of classification of attacks, the CICFlowMeter (3), gathers information of the bidirectional flow. In this work we intend to aggregate more variables to this feature set, considering, for example, other packets received and sent from/to the same IP address (that can target other ports), and generating attack specific features. We hope this kind of feature will enhance the performance of our classification model.

Also, we intend to create a cache policy to store information on IPs, having always as goal to build something with low memory and computational resources consumption. With this we would have the information to generate features mentioned above to the flow classification in real time.

References

[1] Anna-senpai, "Mirai botnet source code." https://github.com/jgamblin/Mirai-Source-Code , 2020. Accessed: 2020-04-15.

[2] I. Sharafaldin, A. Habibi Lashkari, S. Hakak, and A. Ghorbani, "Developing realistic distributed denial of service (ddos) attack dataset and taxonomy," pp. 1–8, 10 2019.

[3] A. Habibi Lashkari, G. Draper Gil, M. Mamun, and A. Ghorbani, "Characterization of tor traffic using time based features," pp. 253–262, 01 2017.

[4] M. Andreoni, O. C. M. B. Duarte, and G. Pujolle, "A monitoring and threat detection system using stream processing as a virtual function for big data," pp. 209–216, 09 2019.