Feature extraction and real time DDoS detection

Alan Barzilay and Caio L. Martinelli

Overview

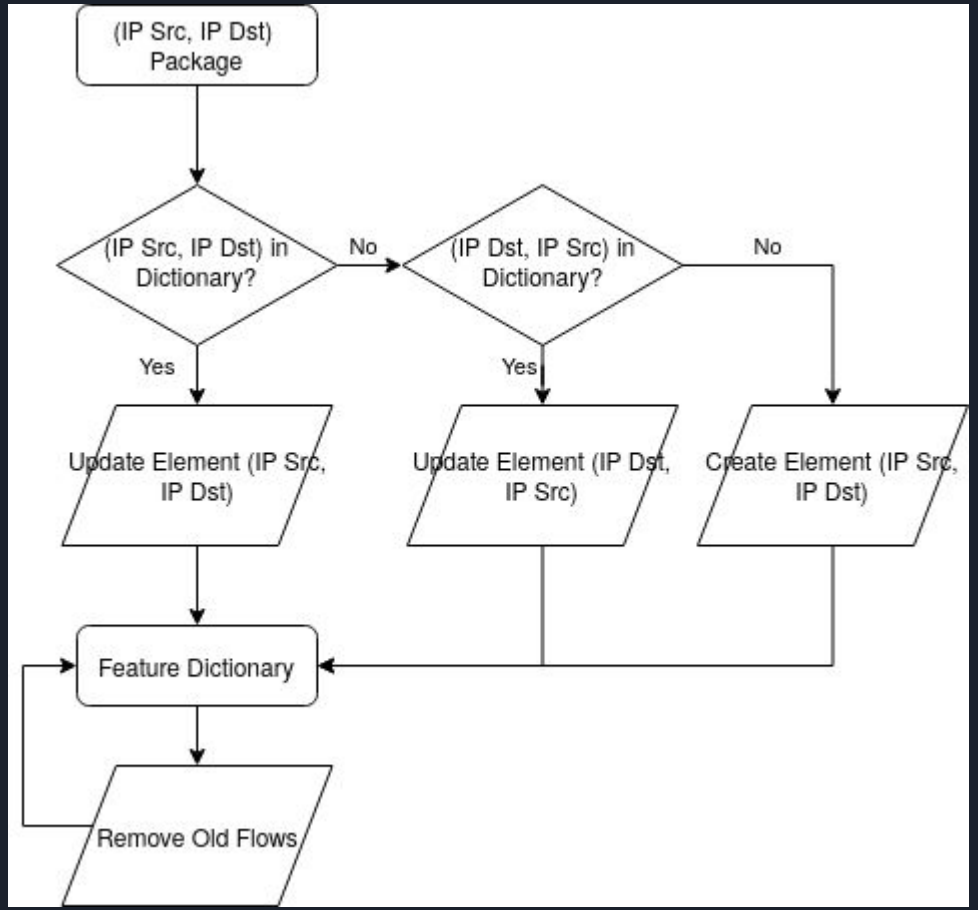
- In order to properly react to a DDoS attack the data stream must be used directly, computation of features must be computationally inexpensive.
- Original dataset has features obtained solely from the 5-tuple flow (IP source, IP destination, port source, port destination, protocol).
- There is literature on features for DDoS attack detection that can’t be obtained restricting the information on the 5-tuple flow (e.g. number of ports that one IP tries to access).
- Counter-measures on DDoS attacks include IP ban, sending CAPTCHAs to the IP source to validate the traffic, limiting number of connections for a single IP, etc.

Proposal

- Create tool to generate features:
  - Store some information of the 5-tuple flux.
  - Store some information of the 2-tuple flux (IP source, IP destination) to generate features based on more than just the bidirectional flow.
  - Store in efficient manner and with “garbage collection”.
  - Readily use stored information to generate features for the flux.
- Evaluate our features against the CICFlowMeter ones.
- Evaluate different models on the data.
- Create a policy to filter packets using those models.

Update of Flows Dictionary

Feature Dictionary:
{ (IP Src, IP Dst): [
Set of source ports,
Set of destination ports,
Set of protocols used,
Number of packets sent,
Number of packets received,
First timestamp,
Last timestamp,
Amount of SYN flags,
...]}



Comparing with CICFlowMeter Features

The CICFlowMeter generates features on the finished flows.

We will generate the features after the last packet is received, so we have comparable features.

Calculate Information Value (IV) and Pearson's correlation with the target.

Classification Models

Use the finished flux to calibrate and evaluate the models. This have two goals:

- More statistics of the features to compare with the CICFlowMeter.
  - Measure of the feature importance in models.
  - Measure of the model performance with/without the features the CICFlowMeter.
- Compare statistics of the models performance with the ones obtained on the dataset paper.

Packet Filter Policy

After choosing the best technique in the last step, use the calibrated models or recalibrate inserting information of the unfinished fluxes.

Use validation data (part of the train data) to create a packet filter policy - probably will be a rule based on IPs. This rule could be used on a firewall such as Linux's iptables.

Conclusion and Future Work

This paper creates a tool that is cheap in computational resources and is specifically designed to detect DDoS attacks.

In the next iteration of this paper we intend to have built and validated the feature tool and to have first results of the models and the packet filter policy.

In future work, we intend to further investigate the computational resources required to keep this system we built working. Also to investigate more mechanisms of counter-measures to take against malicious packages and flows.

Bibliography

1. Anna-senpai,“Mirai botnet source code”,https://github.com/jgamblin/Mirai-Source-Code, 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.
5. J. Li, M. Liu, Z. Xue, X. Fan and X. He, ”RTVD: A Real-Time Volumetric Detection Scheme for DDoS in the Internet of Things,” in IEEE Access, vol. 8, pp. 36191-36201, 2020.
6. Gharaibeh, Manaf, and Christos Papadopoulos. “Darpa-2009 intrusion detection dataset report.” Tech. Rep. (2014).

Questions?

- https://www.overleaf.com/2181452291jhqnvfdhthkc