Feature Engineering for Unknown Web Attacks

Jayesh Gupta, Trevon Helm, Yvonne Kuo, Eric Hope, Pranav Manjunath

Background

- Current intrusion prevention systems utilize web attack signatures to identify malicious behavior.
- Networks are susceptible to zero-day/unknown web attacks, which lack these signatures.
- Analysed 7GB of logs.

Objective

- Engineer new features to detect zero-day attacks.
- Create a scoring system that ranks clients on suspicious activity.

Future Work

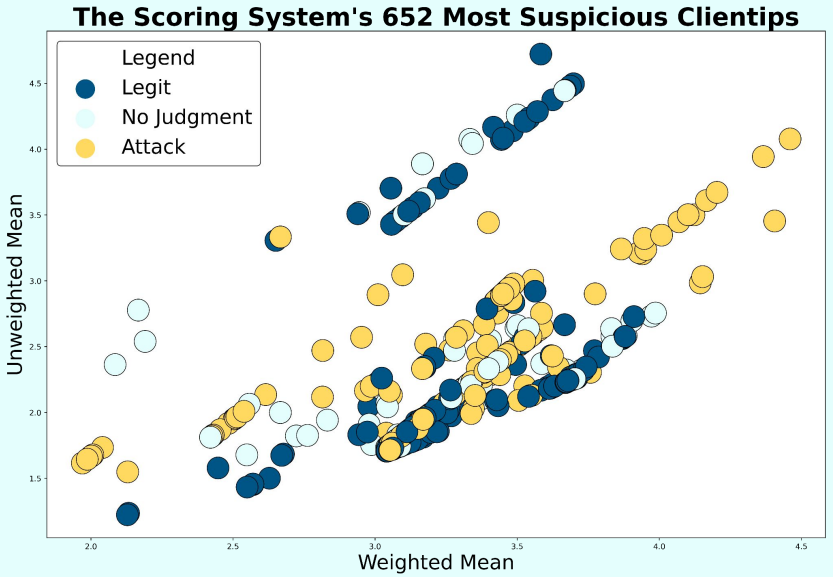
- More EDA to engineer additional features.
- Improve scoring system by minimizing legit bot detection and weighting.

Data

host	web-mgm-02.oit.du...	uri_path	/xmlrpc.php
source	/var/log/httpd/mg...	uri_query	null
clientip	f727e6d830fc2b736...	status	403
ident	-	bytes	212
user	""""	referer	https://craiglab.chem.duke.edu/
_time	2021-05-01T16:58:...	useragent	PHP/6.3.22
method	POST		

Methods

- EDA: to understand the dataset.
- ML techniques: DBSCAN, NLP, KNN to establish normal/irregular behavior.
- Statistical Methods: IQR, Z-scores to examine outliers.



Results

- 6 Detections Created:
  - Blind XSS vulnerabilities
  - Sypex dumpers
  - SEO search abuse
  - Vulnerable adminer.php
  - Commercial fiber router login
  - Wordpress abuse
- 1027 blocked IPs from 81 different countries.

Acknowledgements

- Eric H. (Project Lead)
- Pranav M. (Project Manager)
- Cisco (Sponsors/mentors)
- Iain H.
- Zachary A.