IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

SECURITY MONITORING AND AUDIT PROJECT

G. DOYEN
AND A. JUNIUS
SPRING 2024

PURPOSE

Understand how Intrusion Detection Systems internally operate (partly) beyond their sole installation and configuration

- Implement some basic and standard methods for data analysis
- Follow a state-of-the-art methodology to implement and evaluate several anomaly detectors
- Address a concrete use-case focusing on botnet detection in a real dataset

CONTEXT

The CTU-13 Dataset

The CTU-13 is a dataset of botnet traffic that was captured in the CTU University, Czech Republic, in 2011.

The goal of the dataset was to have a large capture of real botnet traffic mixed with normal traffic and background traffic

The CTU-13 dataset consists in thirteen captures (called scenarios) of different botnet samples

On each scenario the authors executed a specific malware, which used several protocols and performed different actions

The dataset is available here: https://www.stratosphereips.org/datasets-ctu13

CONTEXT

The CTU-13 Dataset

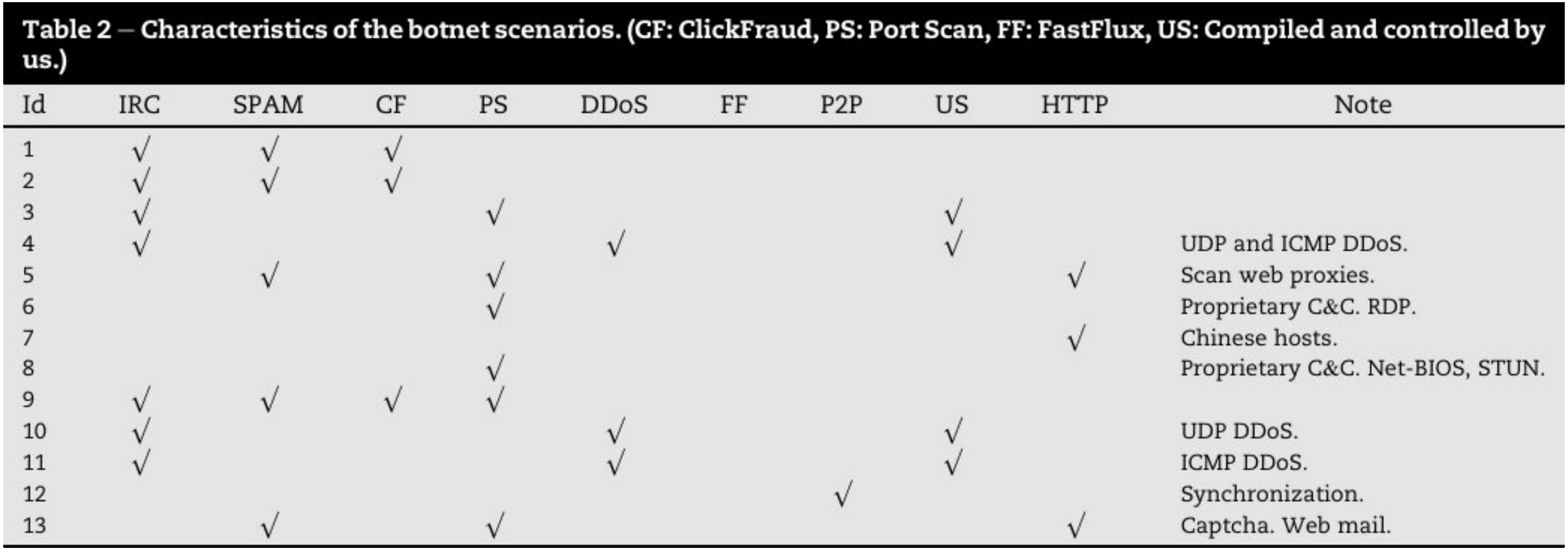

Table 2 – Characteristics of the botnet scenarios. (CF: ClickFraud, PS: Port Scan, FF: FastFlux, US: Compiled and controlled by us.)

Id	IRC	SPAM	CF	PS	DDoS	FF	P2P	US	HTTP	Note
1	√	√	√							
2	√	√	√							
3	√			√				√		
4	√				√			√		UDP and ICMP DDoS.
5		√		√					√	Scan web proxies.
6				√						Proprietary C&C. RDP.
7									√	Chinese hosts.
8				√						Proprietary C&C. Net-BIOS, STUN.
9	√	√	√	√						
10	√				√			√		UDP DDoS.
11	√				√			√		ICMP DDoS.
12							√			Synchronization.
13		√		√					√	Captcha. Web mail.

A GUIDELINE FOR ALL THE STEPS TO ACHIEVE

“An empirical comparison of botnet detection methods”

S. García, M. Grill, J. Stiborek, A. Zunino. An empirical comparison of botnet detection methods. In Computers & Security, Vol. 45, Pages 100-123, ISSN 0167-4048, Elsevier 2014.

- Available at https://www.sciencedirect.com/science/article/pii/S0167404814000923

A full exploitation of the dataset that details

- How the dataset has been captured, how data are structured
- Three different Intrusion Detection System strategies to detect botnets
- How the performance of the IDS can be evaluated (standard performance metrics, dataset split between training and testing)
- It acts as the main guideline to follow to implement the project

THE IDS TO IMPLEMENT AND EVALUATE

Within The Cooperative Adaptive Mechanism for NEtwork Protection (CAMNEP)

Select an anomaly detection approach among the seven proposed in section 3.2 and implement it

- ► Carefully set the threshold value of the anomaly score

Evaluation

Select one of the five testing scenario (and the related training and cross-validation datasets)

- ► Implement the standard performance metrics computation
- ► Depict, understand and conclude on the performance of the selected approach

PROJECT ORGANIZATION

Week #2 (12/02): Project beginning

Teams of 4 students and careful paper reading and understanding

- Group composition must be provided at the end of the session
- Each anomaly detection approach among the 7 must be covered at least by one group, thus, each groups pre-selects an ordered list of 3 algorithms

Week #8 (29/03): Project report and code

An exploitation report submitted on Moodle for each group

The code, which must be made executable must be provided too

PROJECT EVALUATION

An exploitation report

The report must be organized according to the following outline:

1. Motivation and selection of a dataset subpart
2. Statistical dataset analysis to understand its structure and guide the expected behavior of the subsequent detection algorithm
3. Motivation and selection of the implemented detection approach
   - Pseudo-code of the detection algorithm implementation with a step-by-step explanation
4. Results analysis of the detection performance according to the standard metrics
5. Conclusion
6. Annex: the Python code (which can additionnaly be provided online)

SELECTED TOOLS

Freedom to select the tools within each group, according to the background of members

Python is recommended

- Pandas for basic dataset loading, parsing and processing
- NumPy, SciPy, scikit-learn for data processing
- Matplotlib, scikit-learn for data plot

Relevant information on Python for data processing can be found here: Jake VanderPlas. Python Data Science Handbook. Nov. 2016. O'Reilly Media, Inc.

- Full book available at: https://jakevdp.github.io/PythonDataScienceHandbook/

Many web tutorials on Python data processing

Matlab? R?