Seatwork 10.1 Case Study

Improving RT-IoT2022 Analysis

Pascual | (Moldez) – CPE22S3 | CPE311-Computational Thinking with Python

RT-IoT2022 Dataset

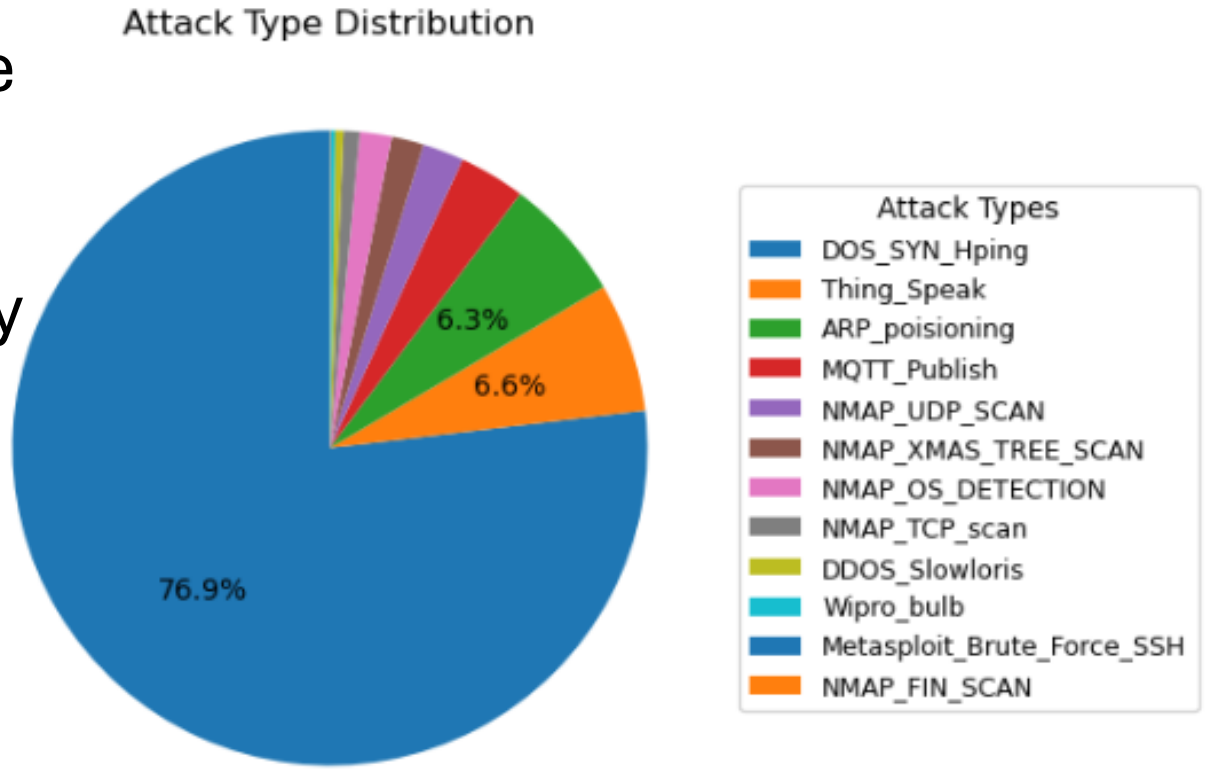
- Primarily used to train Quantized Autoencoder intrusion detection systems (IDS) to detect and identify anomalous attacks in IoT network traffic
- Features / Elements
  - 123117 rows / observations (no duplicates, no missing values)
  - 85 columns (only 15 used for this analysis)

Steps Taken for Analysis

- Preprocessing of Dataset
  - Dropping columns
  - Changing data types to appropriate usage
  - Creating Dataframes for each Attack_type
- Focus on attack types that occurred the least (rare cases)
- Usage of normalized values for the graphs and charts (except for mean and standard deviation)
- Filtering of datasets via IQR to remove outliers
  - Used in finding the mean and standard deviation

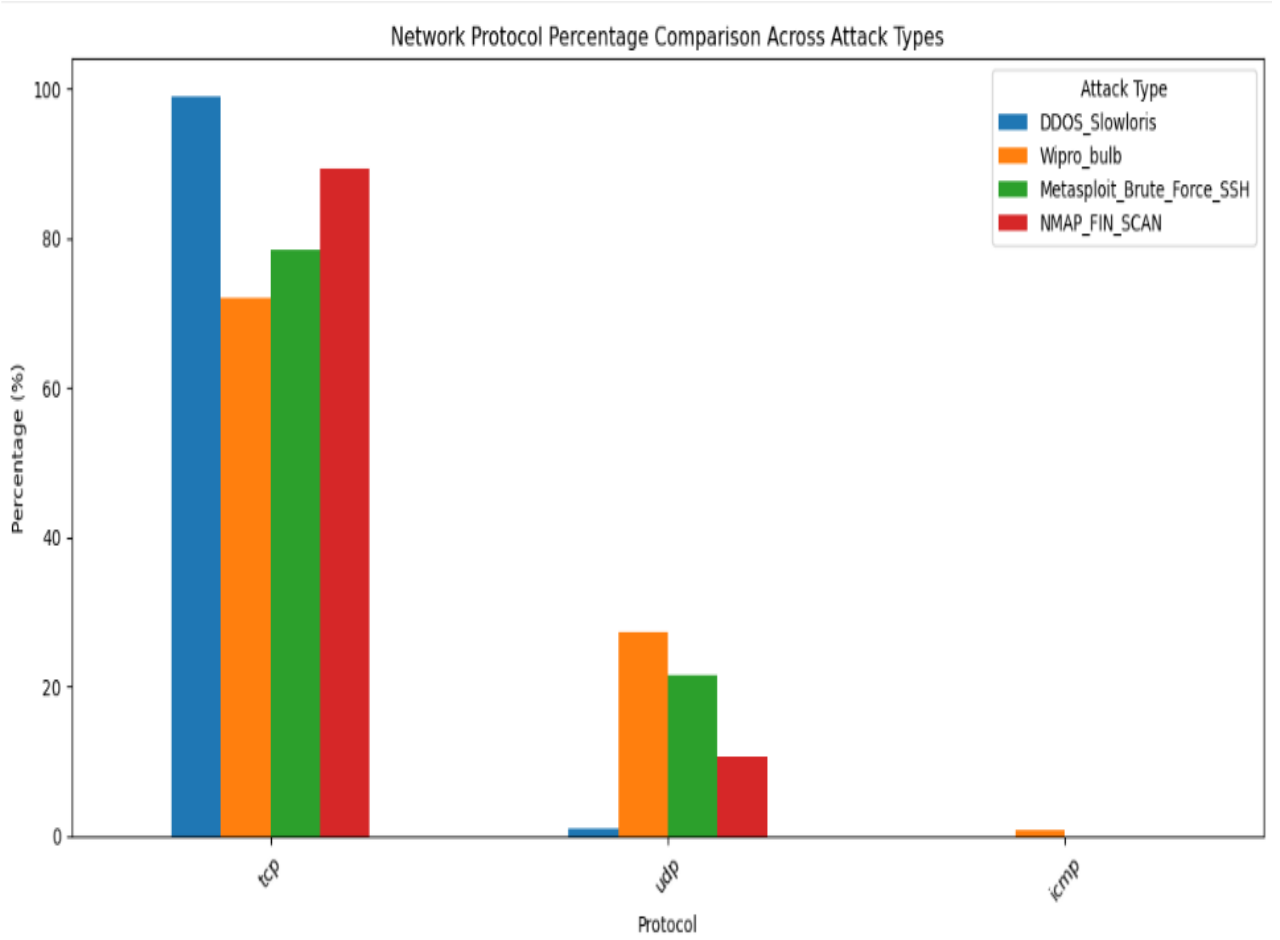
Insights

- The Attack type that occurred the most accounts for ~76% of the observations, while the 4 Attack types that occurred the least only accounts for ~0.69% of the observations.
  - Intrusion detection systems that use this data to train their models have higher chances to perform terribly in minority cases.



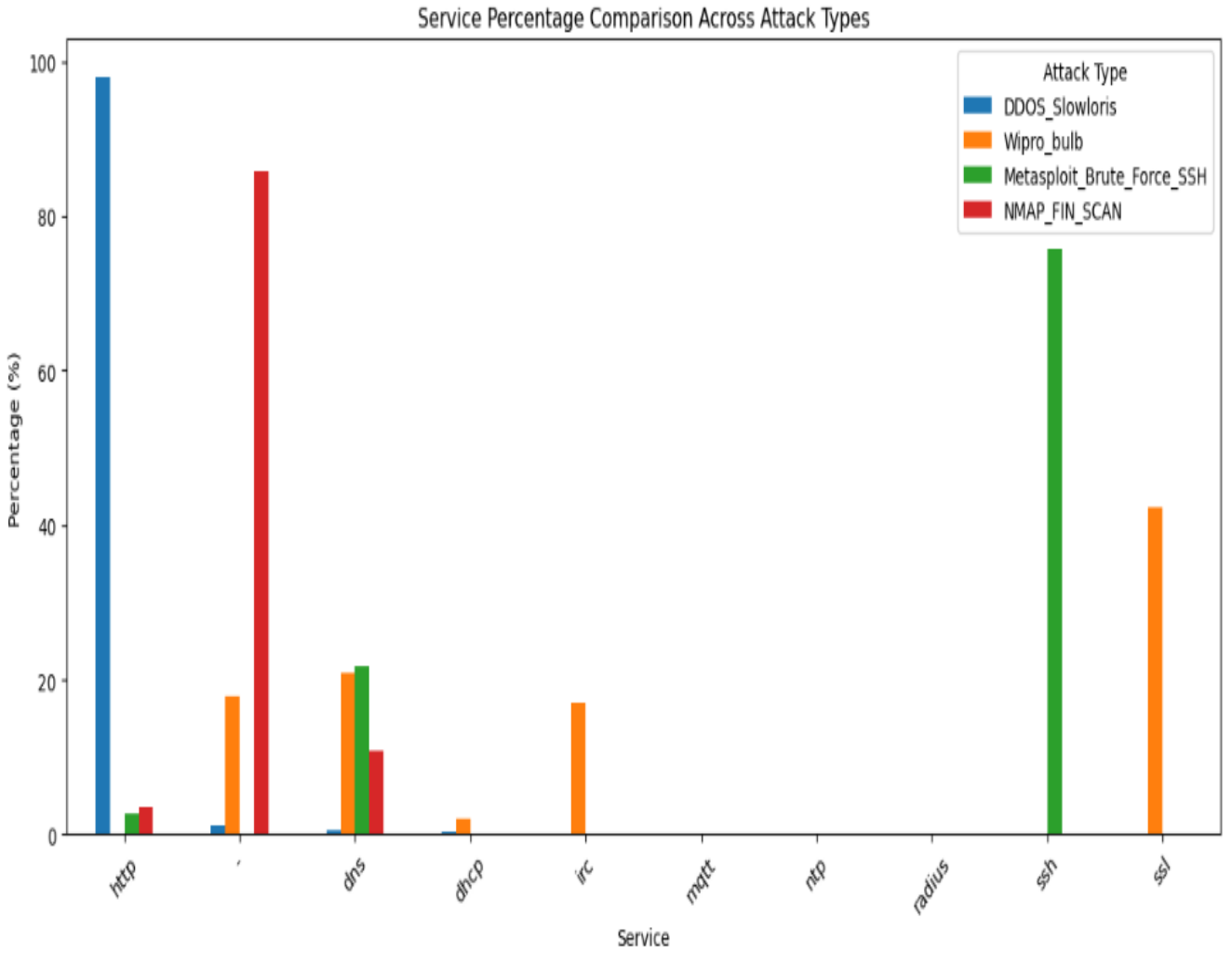
Insights

- Attack types that occurred the least mostly breach or do their damage using the TCP network protocol
- However, attacks using other network protocols should not be ignored too



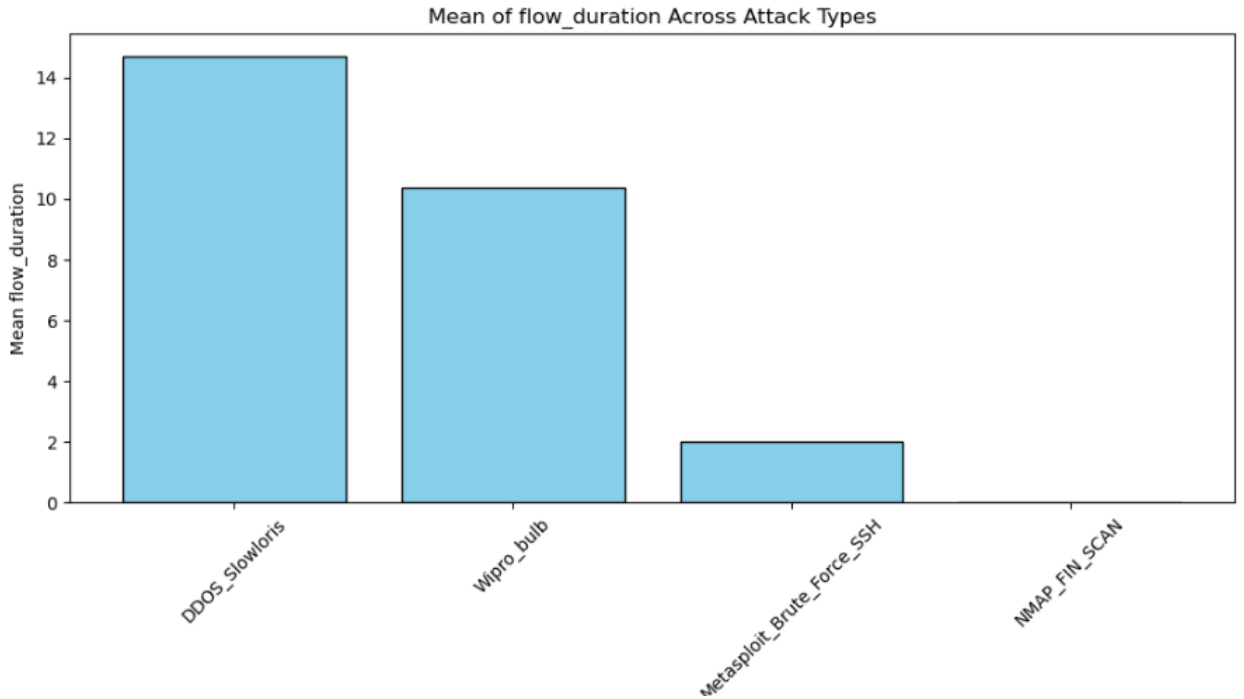
Insights

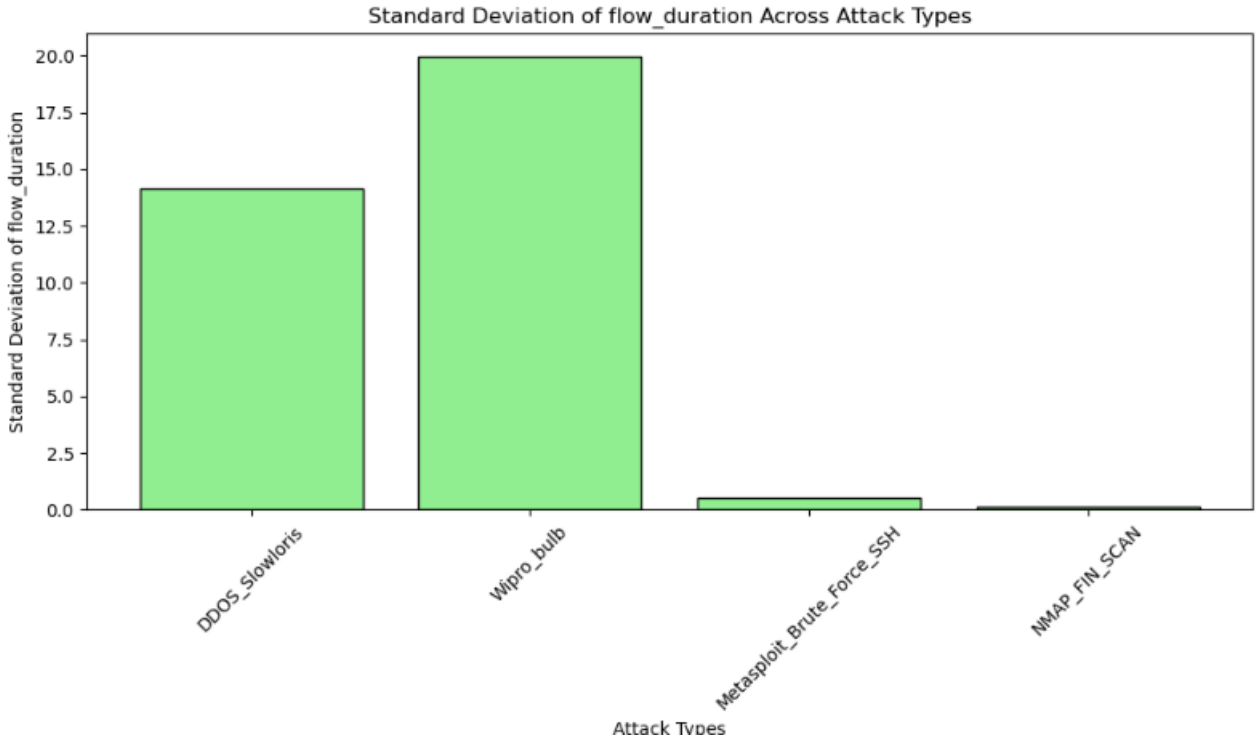
- Some attack types use one service the most
  - DDOS_Slowloris frequently uses http
  - Metasploit_Brute_Force_SSH frequently uses SSH
- However, these attack types can also use other services



Insights

- Flow_duration is consistent amongst all observations for these attack types
  - Brute_Force_SSH
  - NMAP_FIN_SCAN
- Flow_duration varies amongst all observations for these attack types
  - DDOS_Slowloris
  - Wipro_bulb





Recommendations

- More balanced distribution of attack types
- Introduction of attack types that use other network protocols
- Further investigation of attack types that attack through multiple services
  - Service-specific targeting would work for attack types that attack through only one service, but how about others that use multiple services?
- Flow_duration can help in identifying the attack type, based on its value
  - For example, if flow_duration is ~2 units, maybe it’s time to determine if the attack type is Metasploit_Brute_Force_SSH
  - Not applicable for all attacks

Thank you!

References