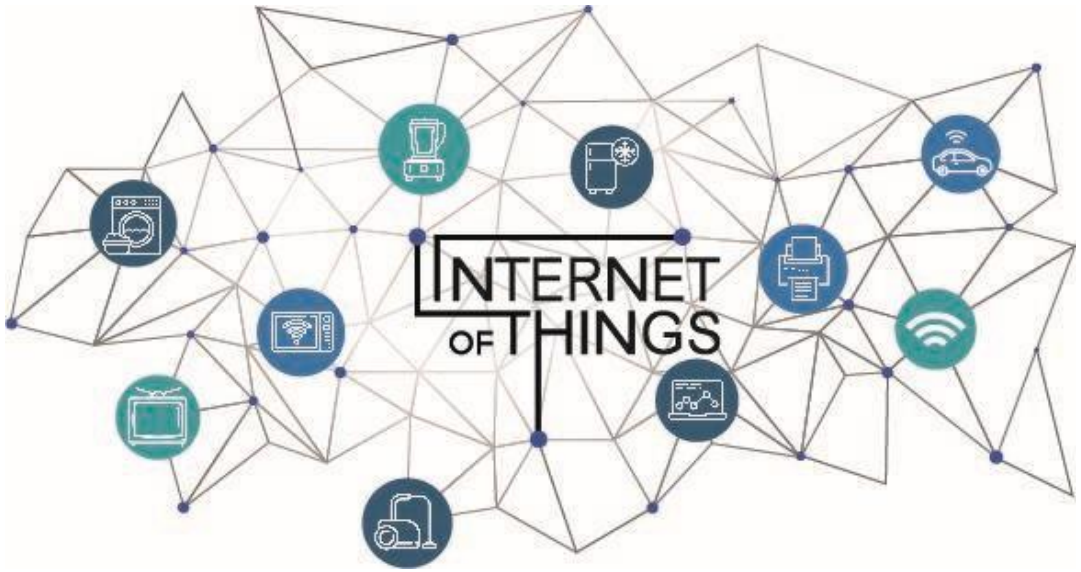
Improving Intrusion Detection in IoT Networks

Silvio Russo, **Isabella Marasco**, Karina Chichifoi, Claudio Zanasi

Department of Computer Science and Engineering, University of Bologna, Italy

Introduction

- **Connectivity between devices** that share data increases the attack surface
- A Network Intrusion Detection Systems (NIDS) is the first component of a complex security system
- Modern NIDS adopt Machine Learning (ML) algorithms to automate the detection process to identify patterns in network traffic
- ML-based NIDS analyze large volumes of traffic, but generate too many false positives



Contributions

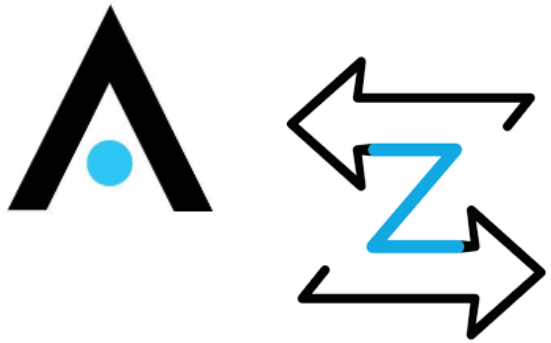
- **Feature extraction** and **selection** → provide the model with a high quality dataset
- **Bayes Point Machine** (BPM) a model for detection
- **11GB dataset** containing benign and malicious samples

Feature Extraction

Feature extraction permits to **extract information** from the data and **removing** irrelevant details and **noise**

We use two network analysis tools:

- Argus
- Zeek



The extracted features are merged into a CSV using the packet timestamp as the key

A total of **50 features** are **extracted**

Feature Selection

Feature selection chooses the most relevant features for classification

The **Mutual Information Classification** algorithm evaluates the mutual information between pairs of features using the entropy of the variables

- Apply **Ward's linkage** method for clustering
- Select the most influential feature per cluster

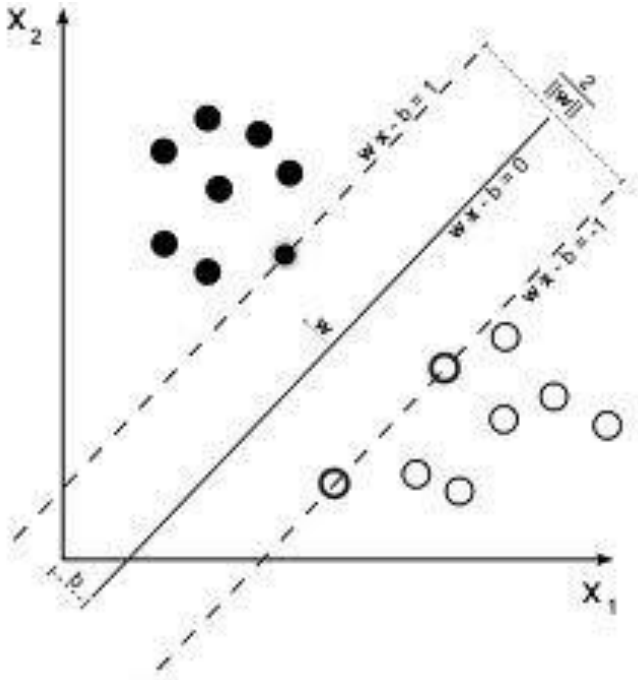
Advantages:

- Mitigate redundancy
- Ensures the preservation of pivotal variables
- Enhancing efficacy and reliability of classification

Bayes Point Machine

Bayes Point Machine is based on Bayesian inference

- **Goal**: Approximate Bayesian posterior of a linear classifier
- **Key feature**: Averages multiple decision boundaries, called “Bayes points” derived from the posterior
- **Benefit**: Improves generalization by combining multiple hypotheses



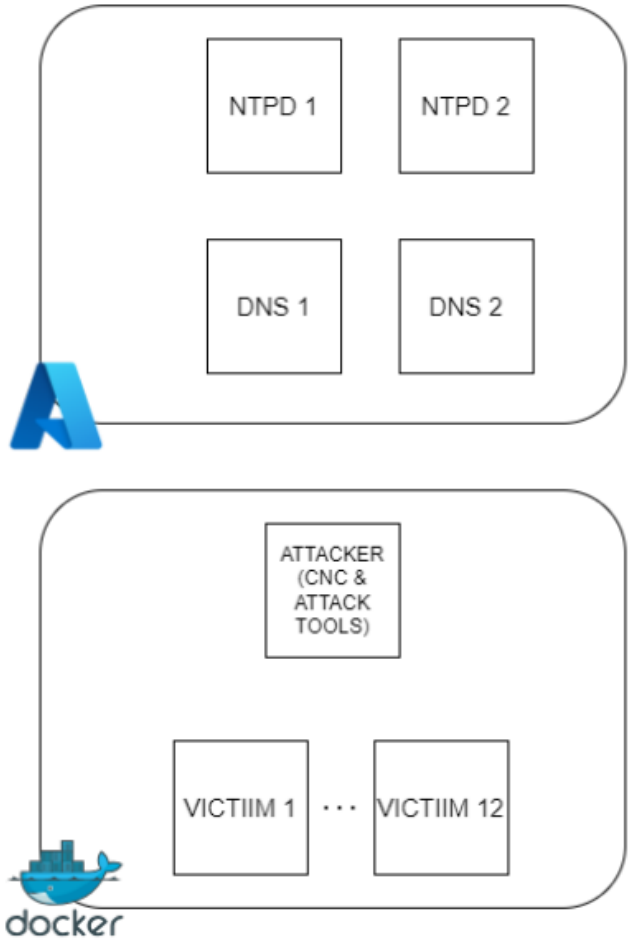
Environment setup

To evaluate the proposed method, we set up an **environment** on **Azure**

The **testbed** was created **based on** the **UNSW-NB15 dataset** for reproducibility reasons

Advantages:

- **Control** over the **attacks** to be tested
- **Prevents** the generation of **new traffic** that could **affect** normal **traffic**



Dataset

Benign traffic:

- Record the traffic of 14 D-Link devices
- HD WiFi Camera
- Wireless N Network Camera
- mydlink Home Smart Plug
- mydlink Home-Connected Home Hub
- mydlink Home Door/Window Sensor

Attacks:

- Web Fuzzing Attack
- Brute-force Attacks
- SSL DoS Attacks
- Remote Code Execution
- Mirai

Experiments Results

To verify the validity of our methodology, we compare the results obtained from different models used as feature extractor:

- Our proposed solution
- Principal Component Analysis (PCA)

Classifier	Accuracy	TP	TN	FP	FN
Bayesian Point Machine	**0.9999**	1273000	1272984	0	16
Logistic Regres-sion	0.9737	1215920	1264249	9251	57580
Random Forest	0.9999	1273497	1273212	288	3
Support Vector Machine	0.9962	1271669	1264746	8754	1831

Classifier	Accuracy	TP	TN	FP	FN
Bayesian Point Machine	0.9840	1260000	1265000	17000	17480
Logistic Regres-sion	0.7164	72609	52985	3015	46732
Random Forest	0.7214	77998	48486	7514	41343
Support Vector Machine	0.7382	75830	53613	2387	43511

Conclusions and Future works

- This research underlines the potential to **focus** not only on machine learning methods, but also on **feature extractors**

- Our proposed methodology combines:
  - Zeek and Argus for feature extraction
  - Bayes Point Machine for anomaly detection

- Future work: Include the online learning mode possible with Bayes Point Machine

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Isabella Marasco

Department of Computer Science and Engineering
University of Bologna, Italy

isabella.marasco4@unibo.it