Institut Mines-Télécom

Machine Learning Based Intrusion Detection System for IoT Network

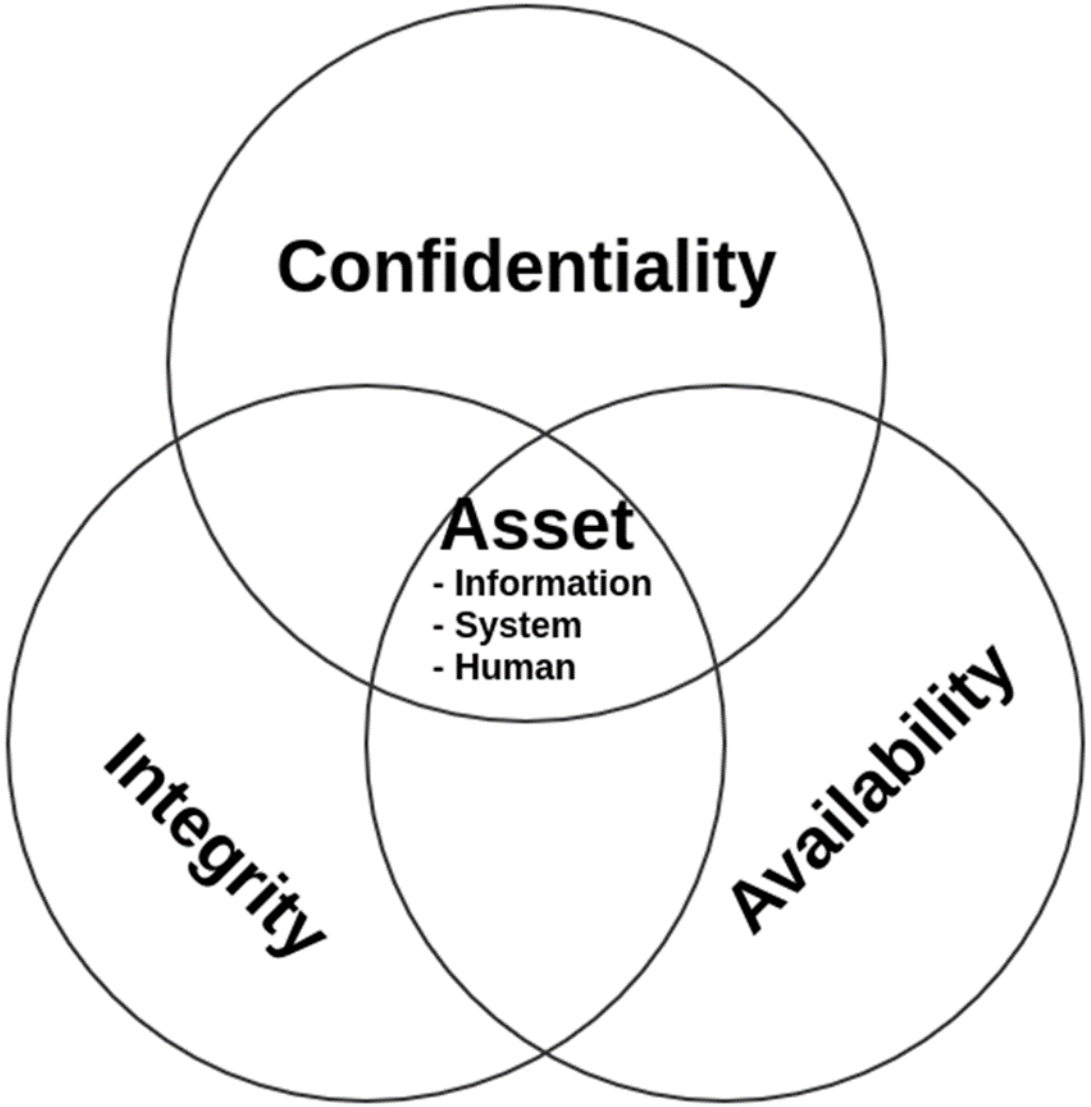
Gregory BLANC & Mustafizur SHAHID
(Télécom SudParis, IMT)

12th July 2018
DATAIA-JST International Symposium on Data Science and AI

OUTLINE

1. Introduction
2. Misuse Detection
3. Past Works
4. Internet of Things
5. Anomaly Detection
6. Data Generation
7. Features Selection
8. Future Works

Institut Mines-Télécom



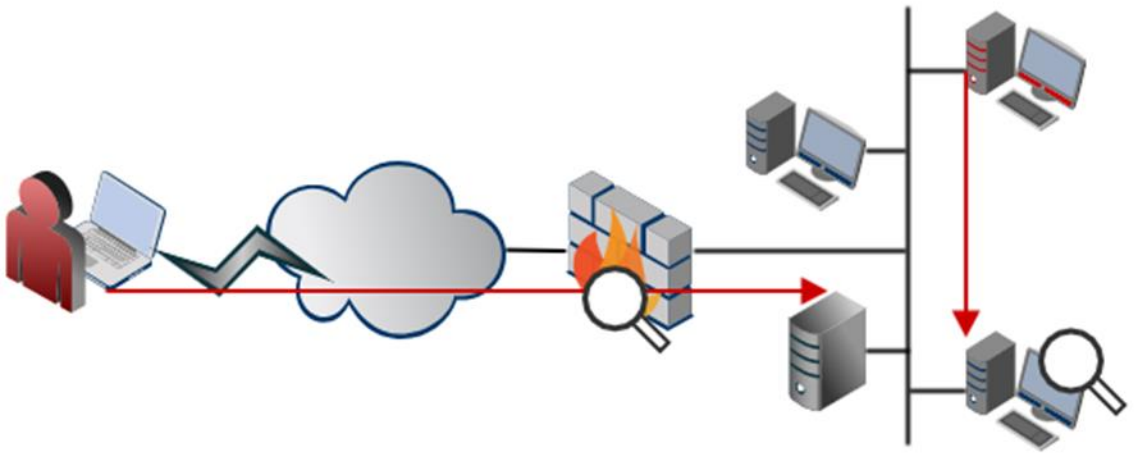
Risks = Impact x Occurrence

Threats

- unauthorized usage
- malicious usage
- alteration
- hijacking
- abuse
...

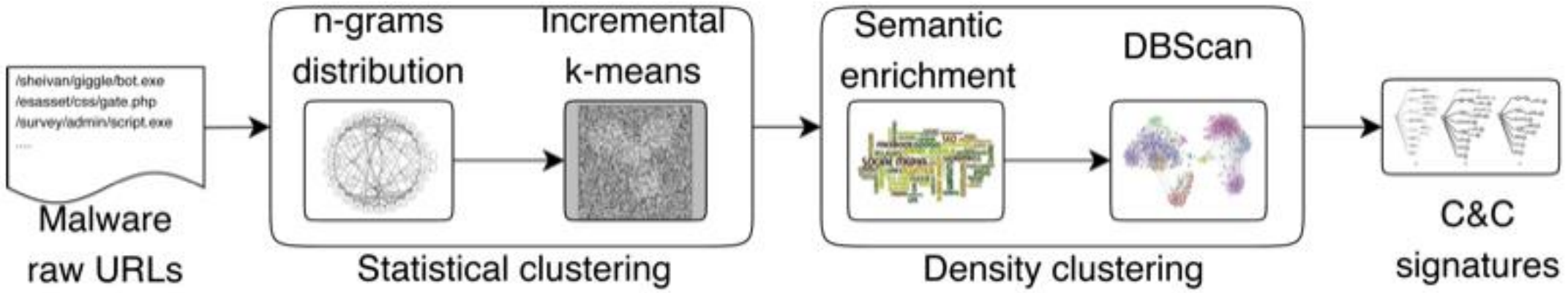
Vulnerabilities

- insufficient physical protection
- insufficient logical protection
- lack of validation of input data
- lack of traceability of input data
- reckless behavior
...



- Huge volume of traffic impacts processing time
  - Filtering
  - Correlation
- Malicious activities require a model
  - Known activities: misuse detection
  - Unknown activities: anomaly detection

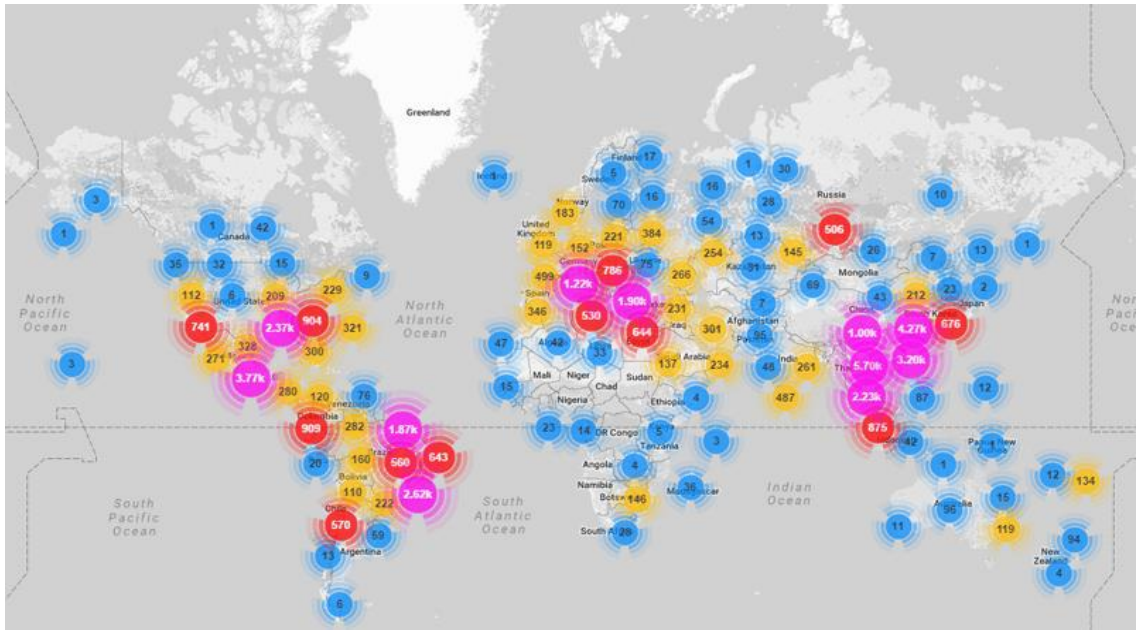
- Mostly used approach: signature-based
- Features: packet headers, flow statistics, TCP connections, etc.
- Relies on data mining and machine learning
- Challenges
  - Lack of datasets (existence, diversity, freshness, sources)
  - Requirement to update the model frequently



- Assumption: botnet machines abuse HTTP to contact their C&C
- Dataset: generated by malware (malicious HTTP requests)
- Goal: generate signatures to detect requests to C&C

Signa-tures	Malware Dataset		
	2011	2012	2013
2011	87%	64%	21%
2012	NA	86%	57%
2013	NA	NA	81%

- Smart home
- Smart grid
- Smart transportation
- Smart industry
- eHealth

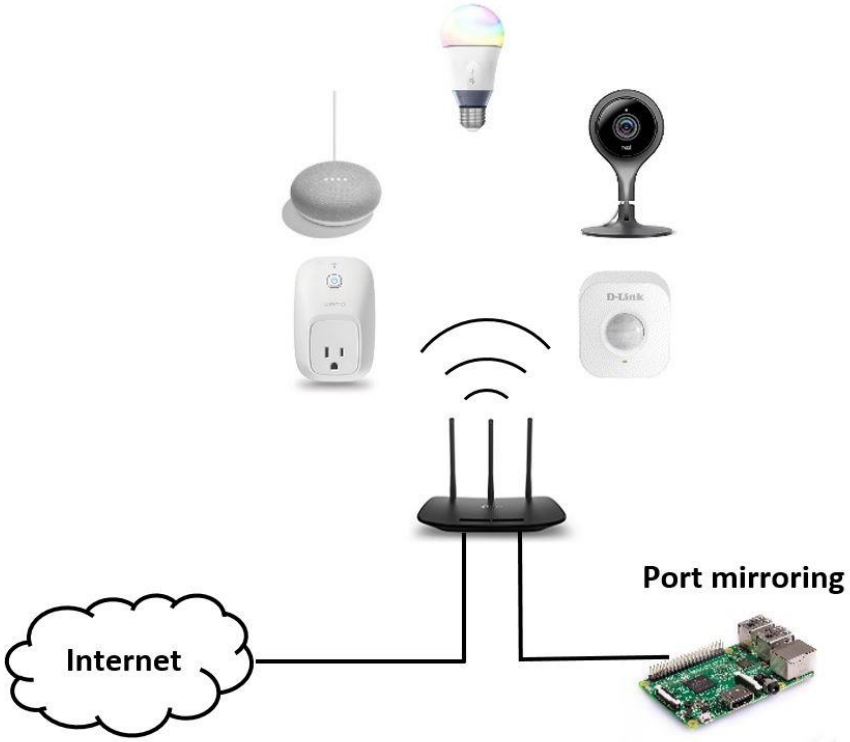


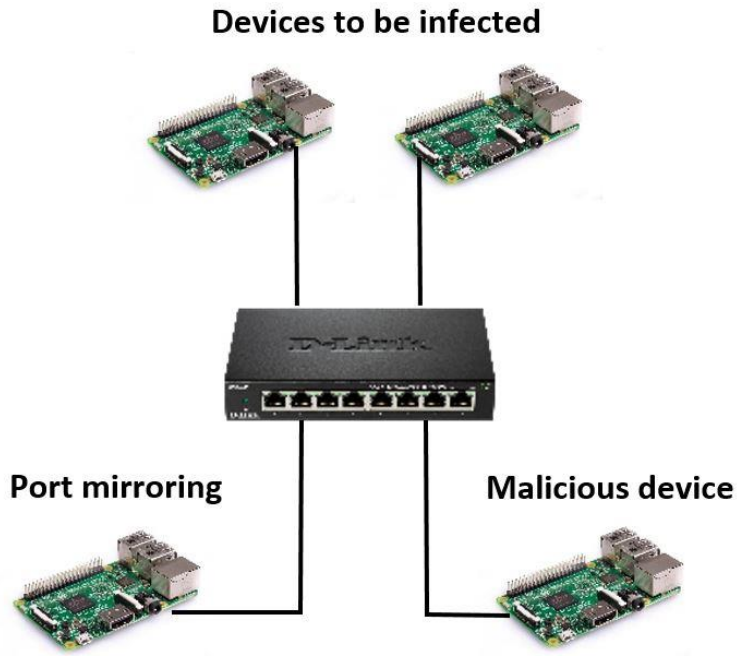
Security issues:

- Short time to market
- Vulnerable default configuration
- Example of Mirai botnet

- *Anomaly detection*: detection of any deviation from legitimate behavior. Define a profile of the legitimate traffic and perform outlier detection. Can detect new types of attacks. High false positive rate.
- IoT devices have very specific purpose, hence, the variation of the observed network traffic behavior is limited.

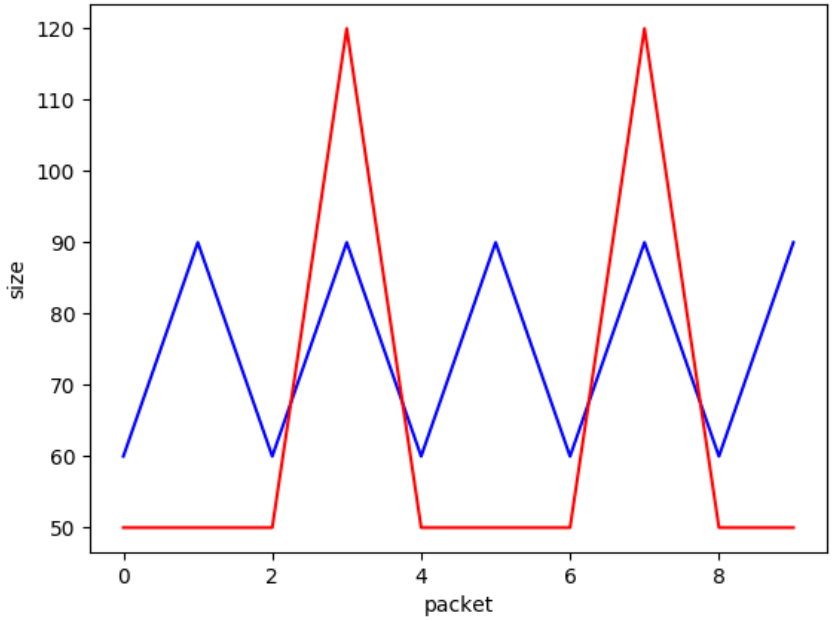
- Limitation of existing datasets (DARPA99, KDD99, NSL-KDD)
- Experimental smart home network for legitimate network traffic generation
- Malicious traffic generation





- Features selection:
  - Avoid features that can introduce bias: IP identifier, addresses, ports number, …
  - Features that can be easily extracted from a practical point of view

- Examples of feature: Individual packet size, total number of packets, TCP flags, inter packet duration, …



- Intrusion detection system for smart home network
- Mimicking legitimate behaviour to evade intrusion detection system in IoT network
- Public IoT network traffic datasets generation for benchmarking purposes

Partners:

- Nara Institute of Science and Technology (NAIST)
- Yokohama National University