IoT Security on the Edge

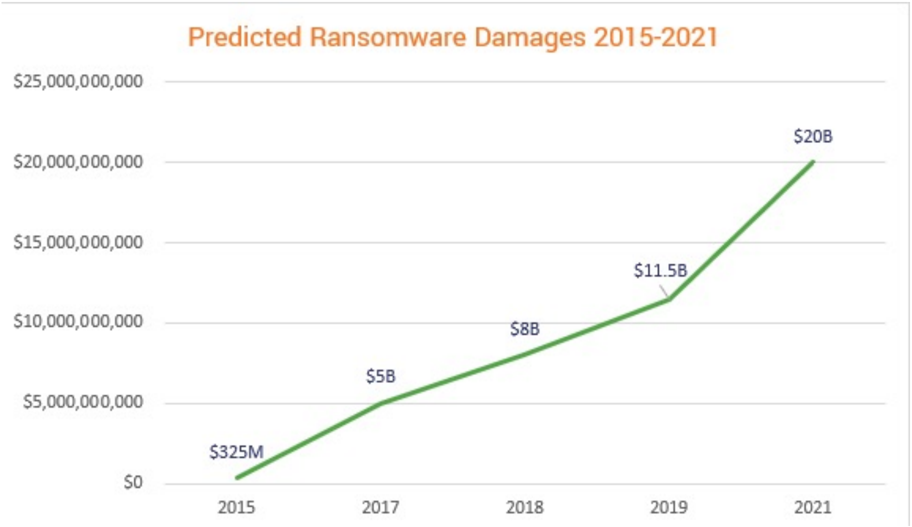
Paul Patras

THE UNIVERSITY of EDINBURGH
informatics
icsa

Cyber-attacks on the rise, more than ever

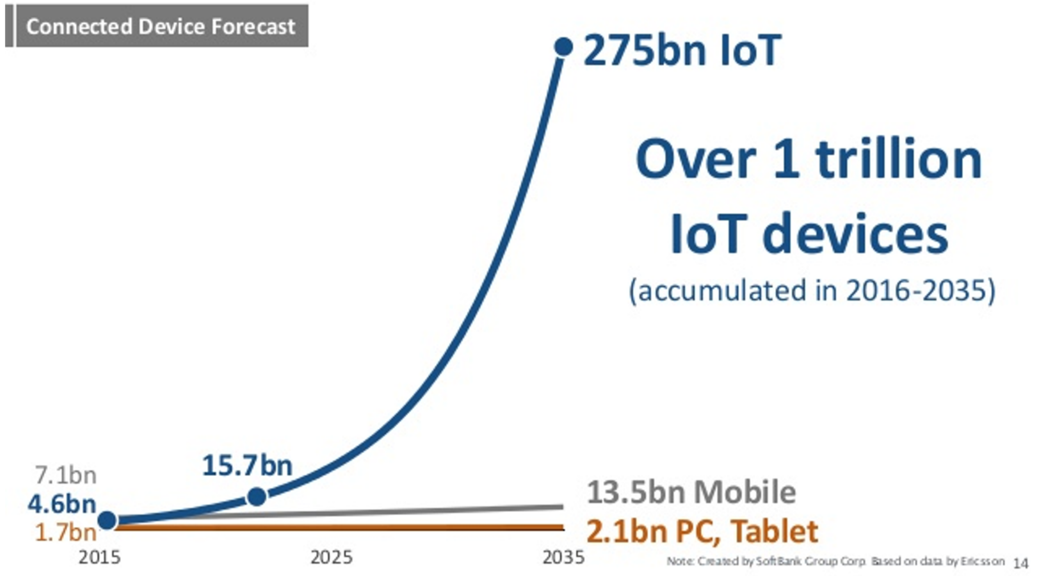
- Cybercrime to cost the world **$10.5 Trillion** annually by **2025** (Cybersecurity Ventures)
- **1.14 billion** malware instances registered by the end of **2020** (AV-TEST)
- Number of DDoS attacks worldwide to hit **15.4 million** by **2023** (Cisco)

Ransomware cost



Cyber-attacks on the rise, more than ever

- Cybercrime to cost the world **$10.5 Trillion** annually by **2025** (Cybersecurity Ventures)
- **1.14 billion** malware instances registered by the end of **2020** (AV-TEST)
- Number of DDoS attacks worldwide to hit **15.4 million** by **2023** (Cisco)
- **1 trillion** connected devices expected by **2035** (ARM)



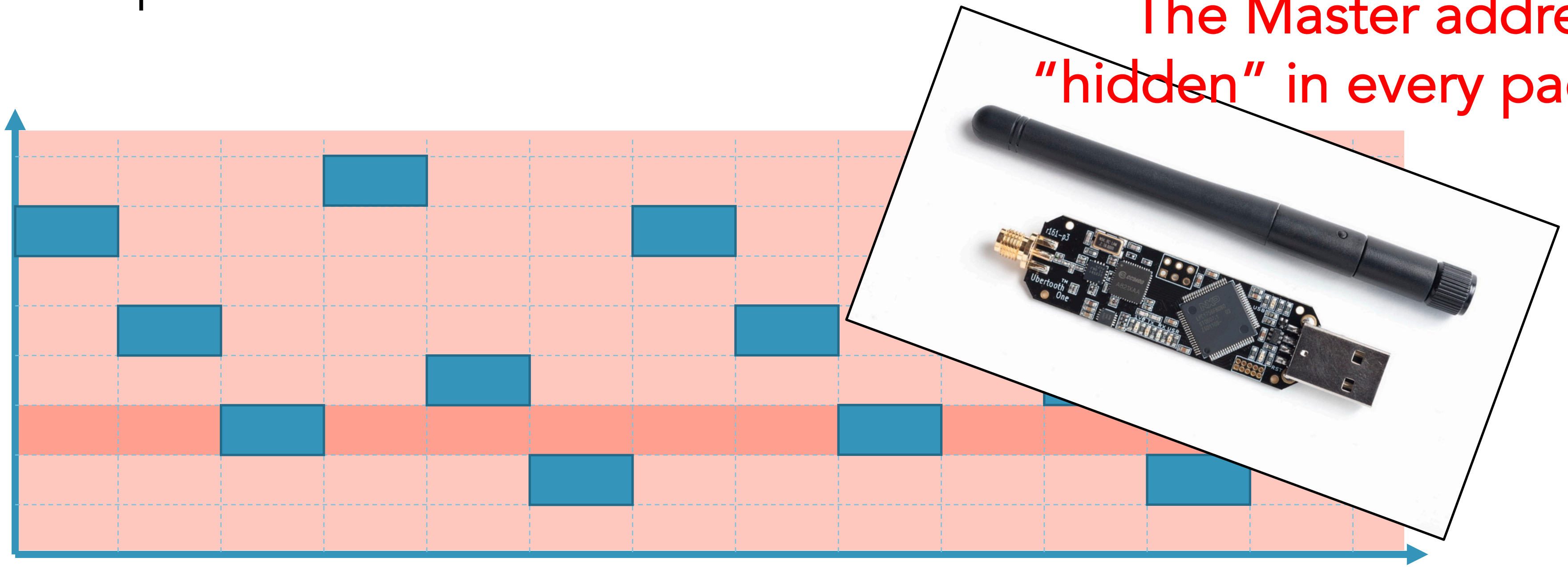
Bluetooth reborn with IoT

- 5 billion Bluetooth devices to be shipped in 2021 (Statista)
- Bluetooth BR/EDR (or Bluetooth Classic) widespread

Connections are hard to sniff

Frequency Hopping

- Pseudo-random hopping across 79 channels
- 1600 hop/s



De-anonymizing Bluetooth Devices

- LAP present in clear in every packet
- Two quantities missing

Master's Clock (6 bit)
Master's UAP (8 bit)

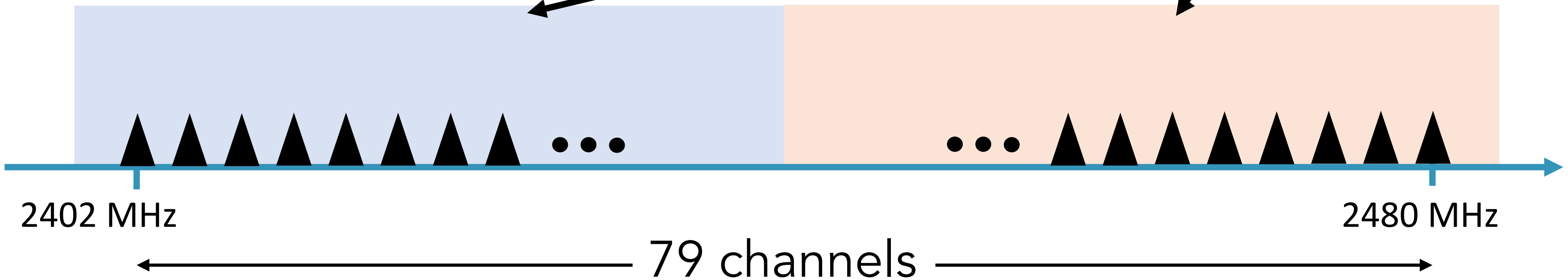
→ 2^{14} possible pairs

Bruteforcing all possible Clock + UAP pairs is feasible!

Building a full-band Bluetooth sniffer

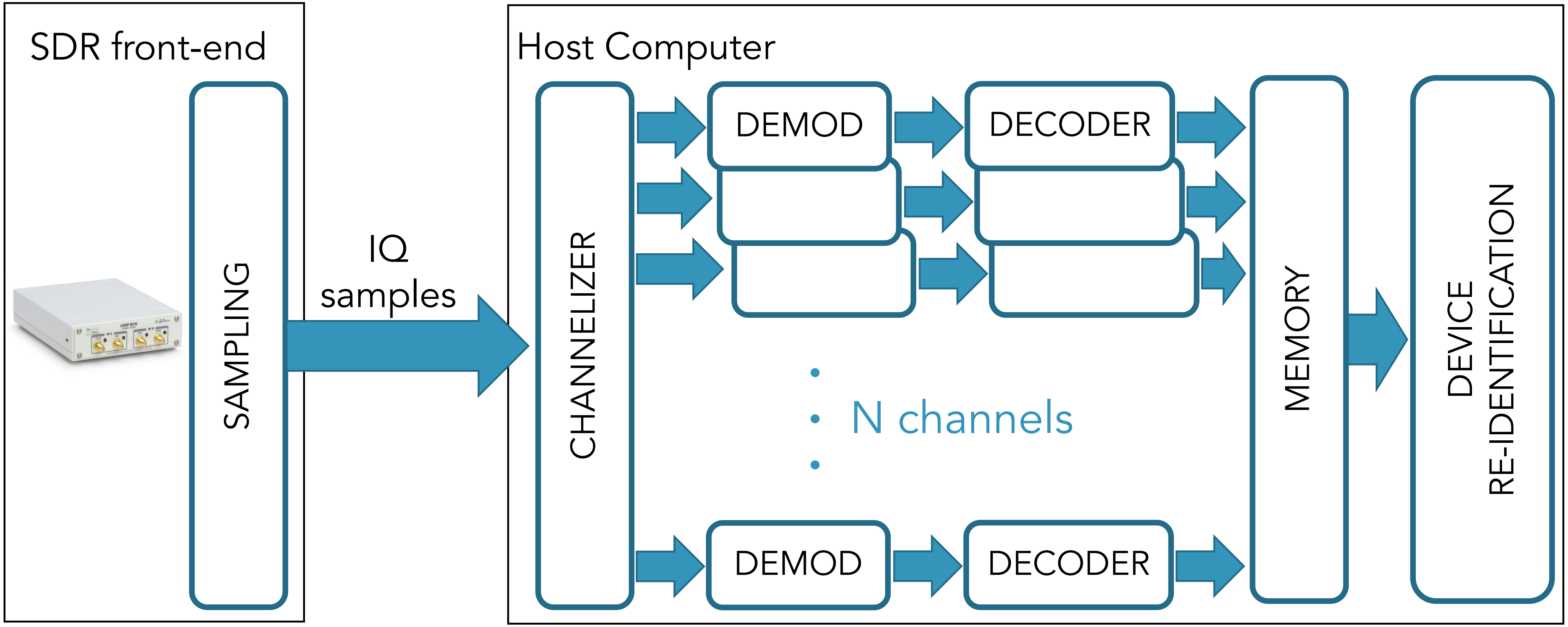
Boards have to be synchronized!

Separate and process narrow-band channels



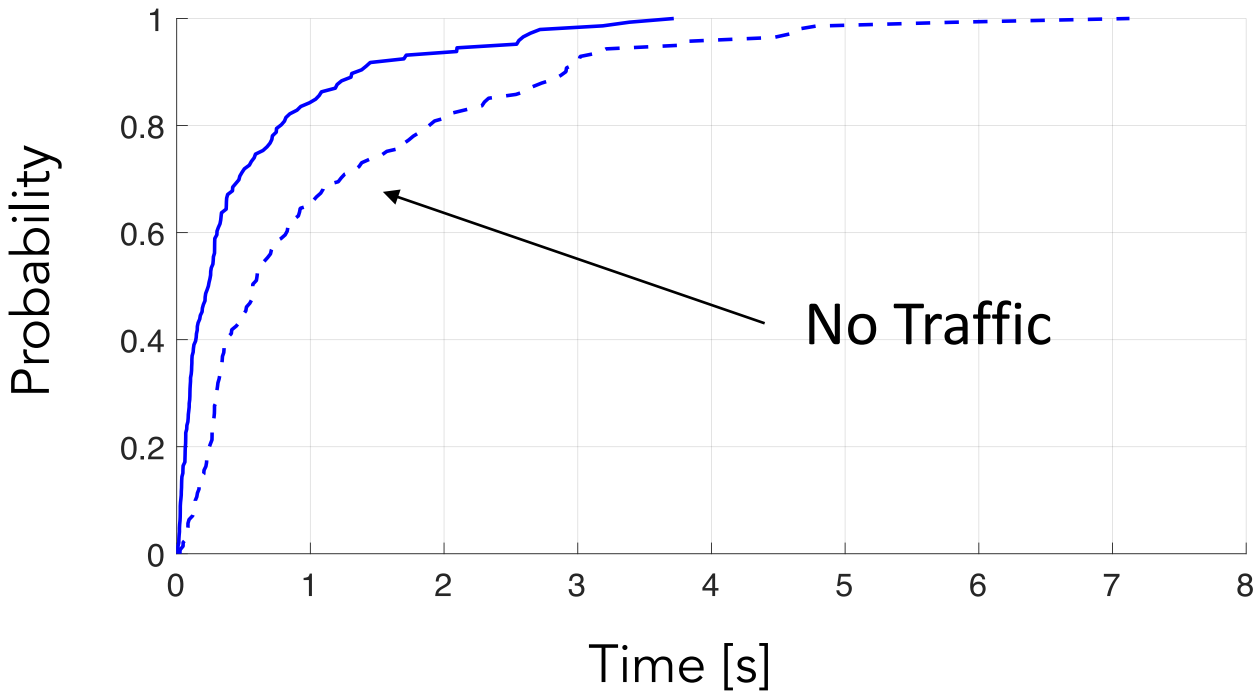
M. Cominelli, F. Gringoli, M. Lind, P. Patras and G. Noubir, "Even Black Cats Cannot Stay Hidden in the Dark: Full-band De-anonymization of Bluetooth Classic Devices," IEEE S&P 2020.

SDR Architecture

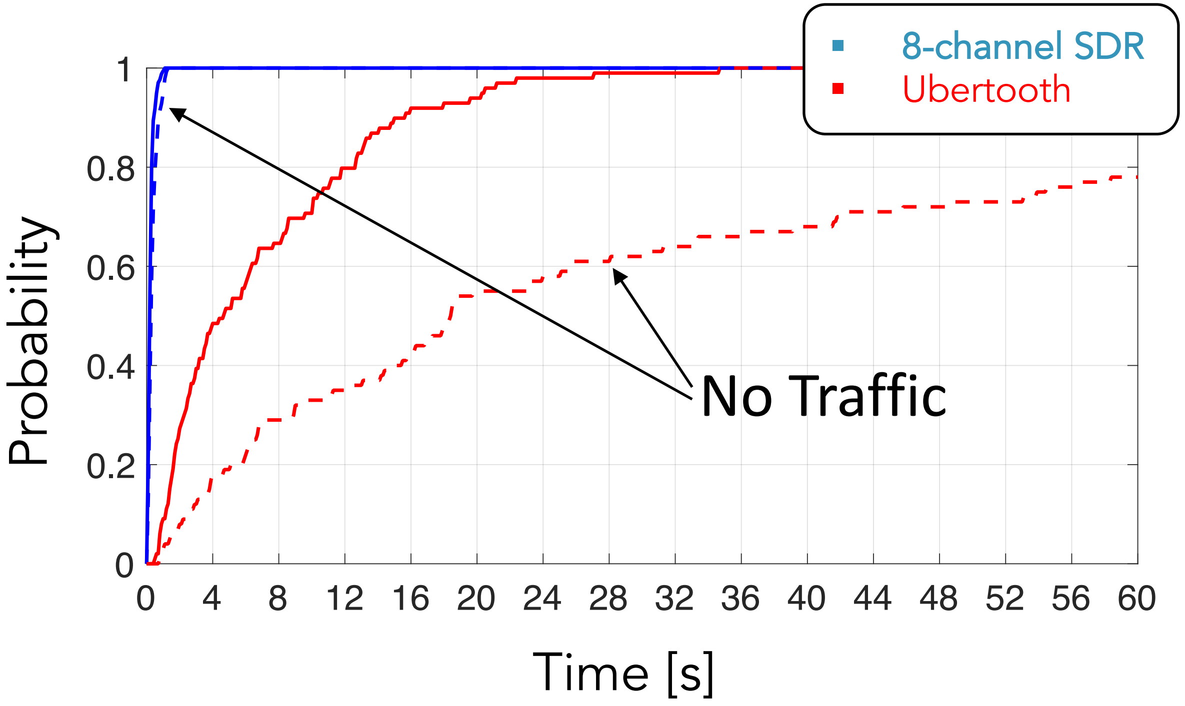


Performance

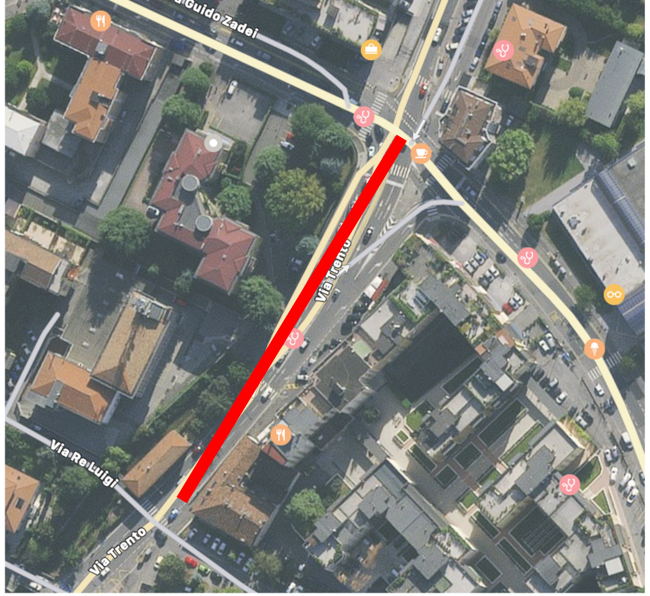
- Majority of 25 connections detected in <1 second



- Car audio system detected orders of magnitude faster than Ubertooth

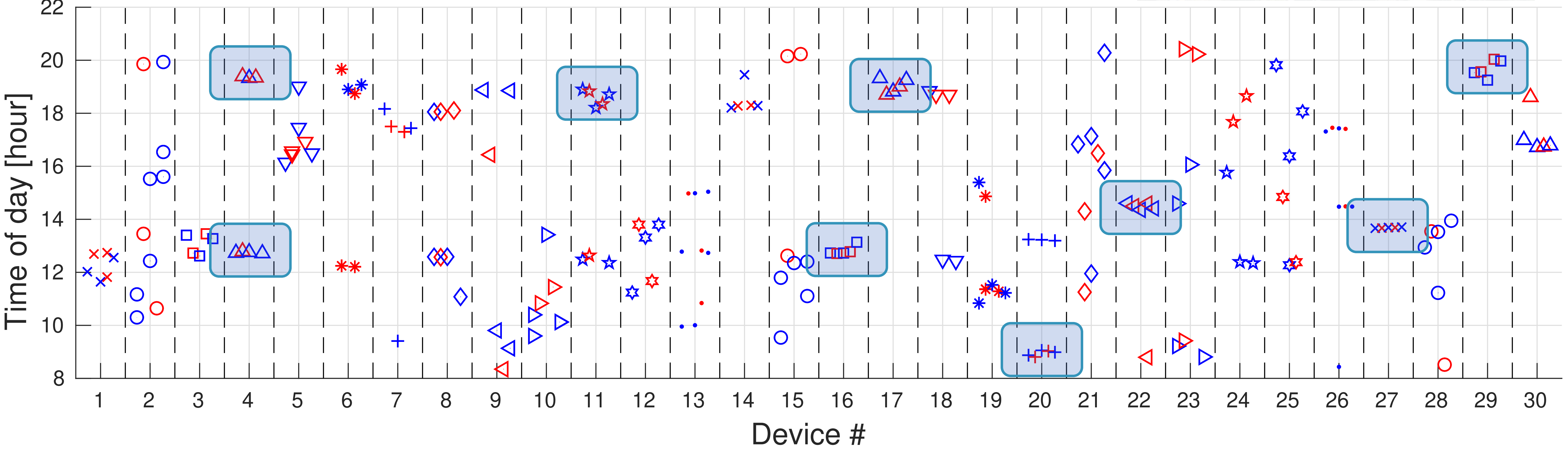


Tracking commute patterns



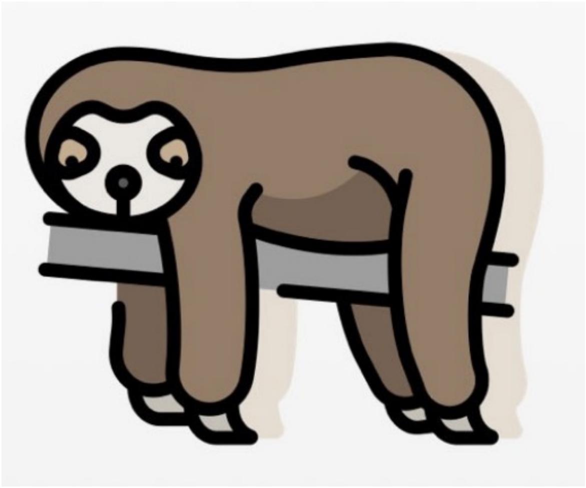
Monitoring traffic at a road junction

5 working days



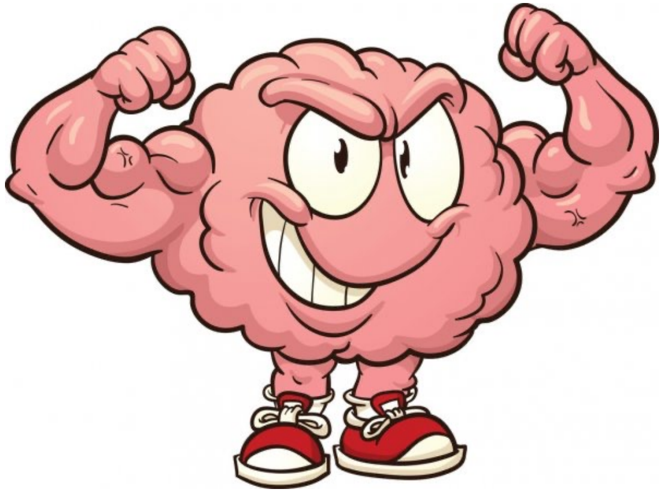
AI/ML-based NIDS solutions getting traction

Rule-/signature-based detection



- Too many false positives
- Significant ongoing maintenance
- Cannot detect unknow attacks
- Etc.

Deep learning approaches



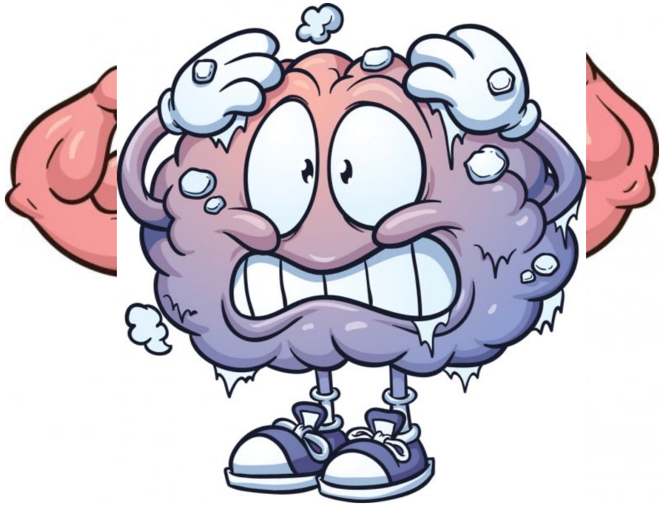
- Easier to detect illicit activity hidden in data traffic
- No need to look at every packet
- Should have decent generalization abilities

AI/ML-based NIDS solutions getting traction

Rule-/signature-based detection

- Too many false positives
- Significant ongoing maintenance
- Cannot detect unknow attacks
- Etc.

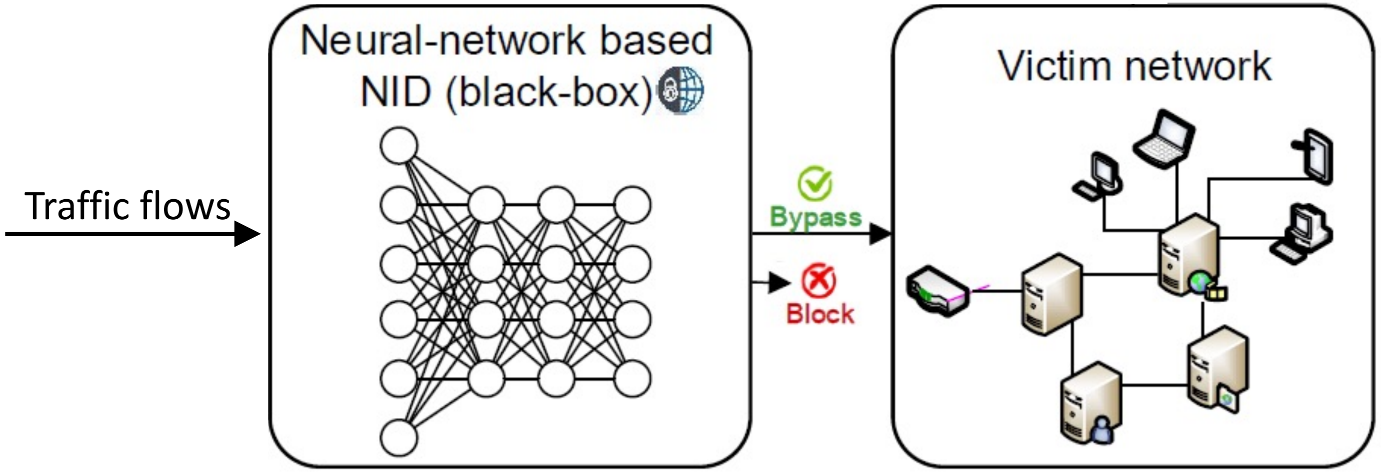
Deep learning approaches



- Easier to detect illicit activity hidden in data traffic
- No need to look at every packet
- Should have decent generalization abilities

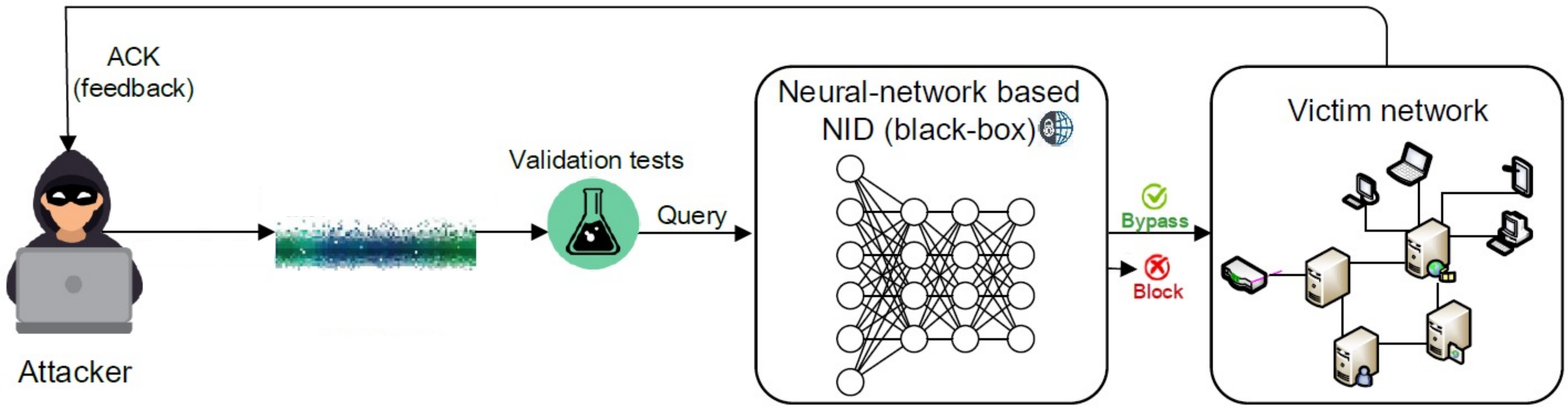
Question: Is DL reliable for intrusion detection?

Threat model



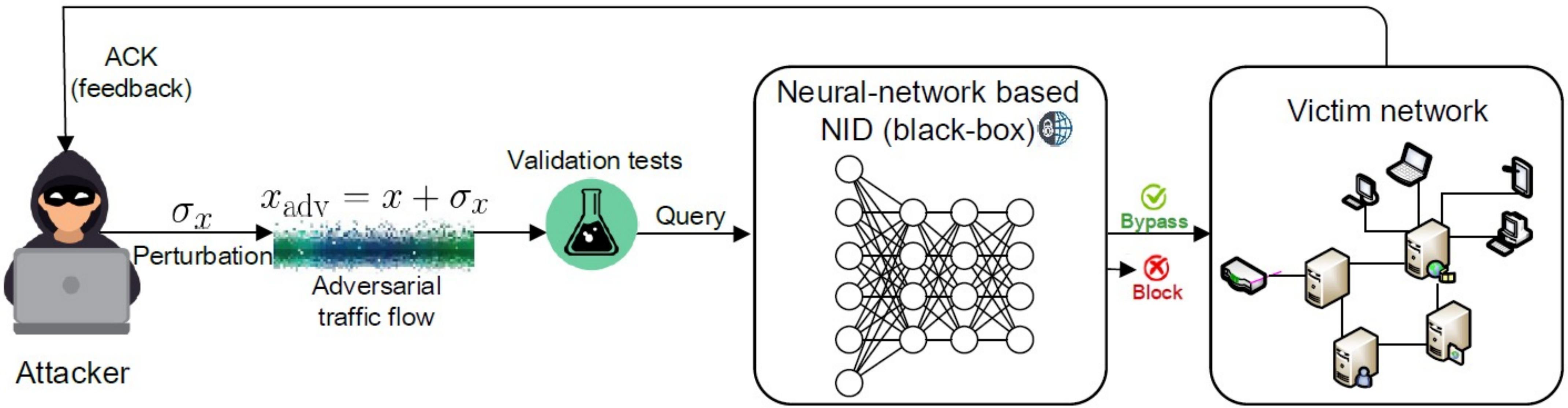
Network intrusion detection system (NIDS)

Threat model



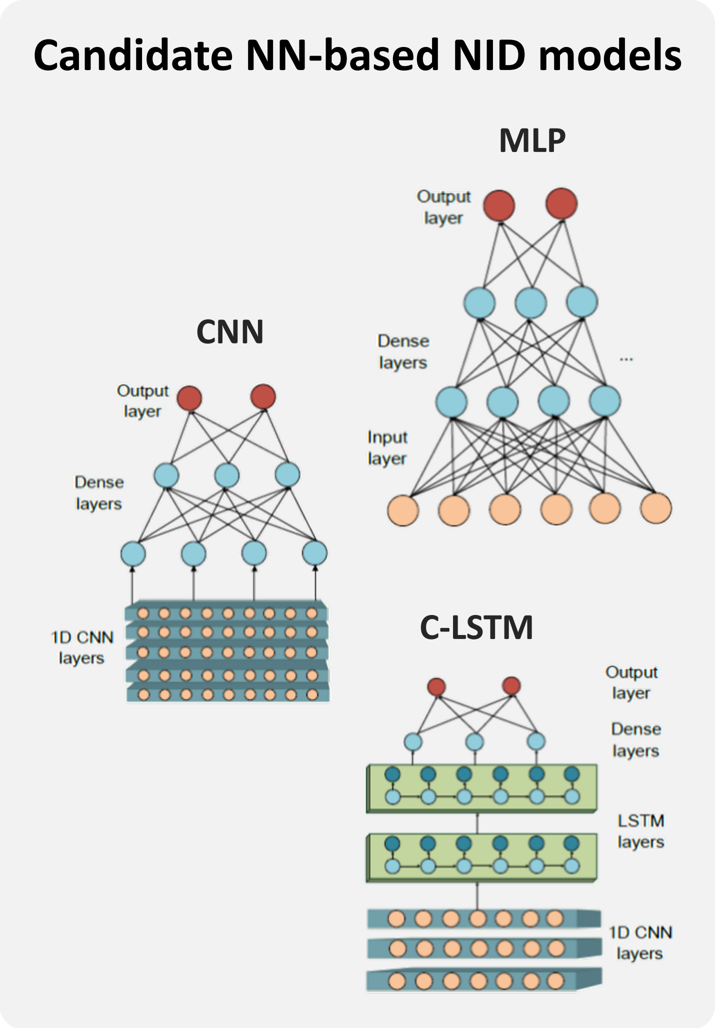
Network intrusion detection system (NIDS)

Threat model



Network intrusion detection system (NIDS)

Tiki-Taka: Adversarial Attacks and Defenses against Them



C. Zhang, X. Costa-Perez, P. Patras, "Tiki-Taka: Attacking and Defending Deep Learning-based Intrusion Detection Systems", ACM CCSW 2020.

Tiki-Taka: Adversarial Attacks and Defenses against Them

Adversarial attacks

NES

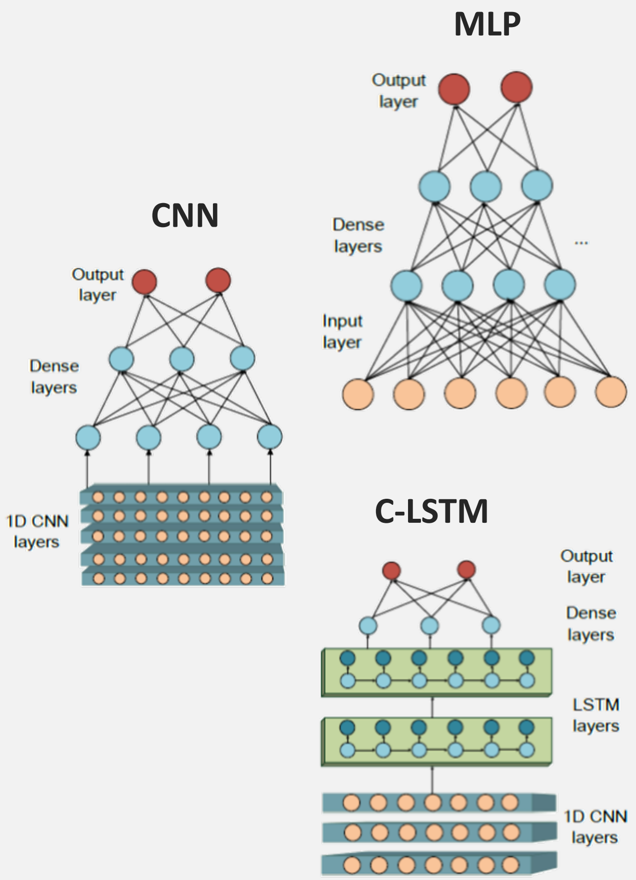
Boundary

Pointwise

HopSkipJumpAttack

Opt-Attack

Candidate NN-based NID models



C. Zhang, X. Costa-Perez, P. Patras, “Tiki-Taka: Attacking and Defending Deep Learning-based Intrusion Detection Systems”, ACM CCSW 2020.

Tiki-Taka: Adversarial Attacks and Defenses against Them

Adversarial attacks

NES

Boundary

Pointwise

HopSkipJumpAttack

Opt-Attack

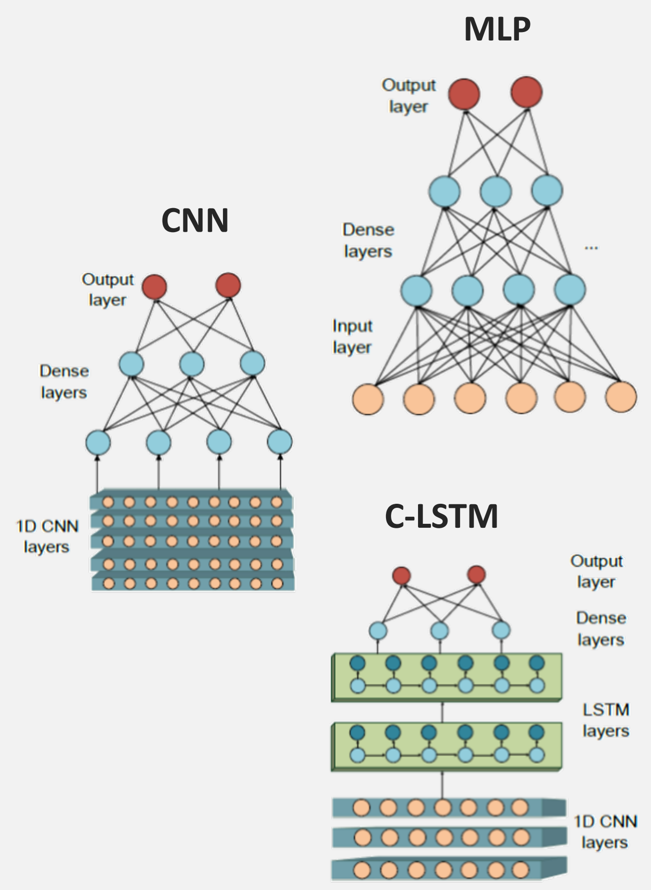
Defense Mechanisms

Model Voting Ensembling

Ensemble Adversarial Training

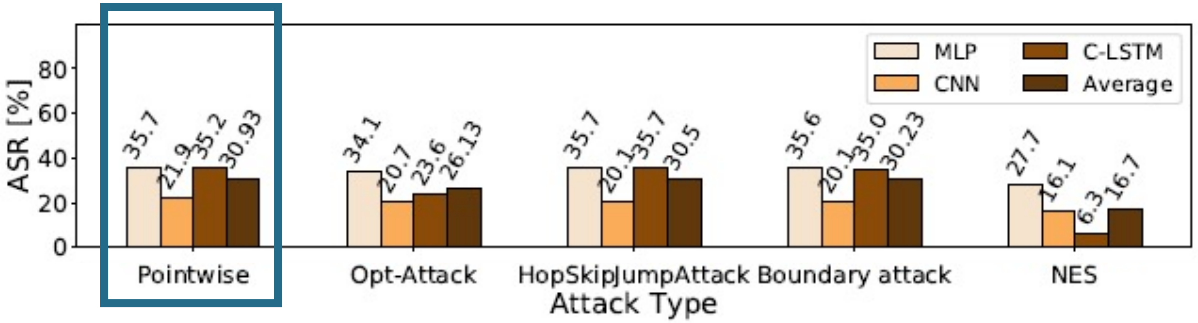
Adversarial Query Detetion

Candidate NN-based NID models



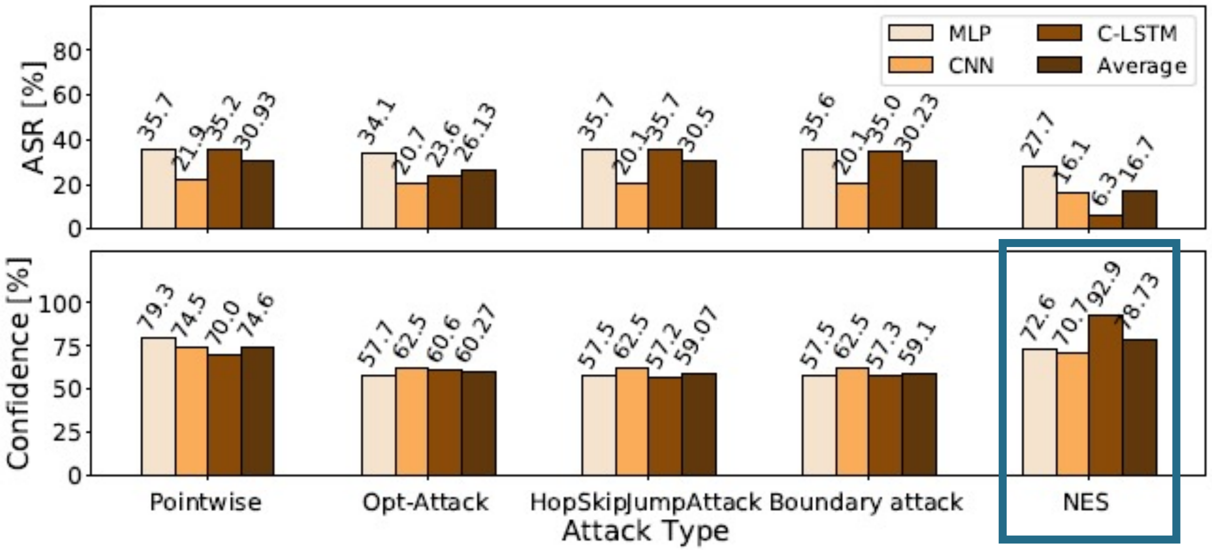
C. Zhang, X. Costa-Perez, P. Patras, "Tiki-Taka: Attacking and Defending Deep Learning-based Intrusion Detection Systems", ACM CCSW 2020.

Adversarial attack performance



- **Average attack success rates up to 35.7%**

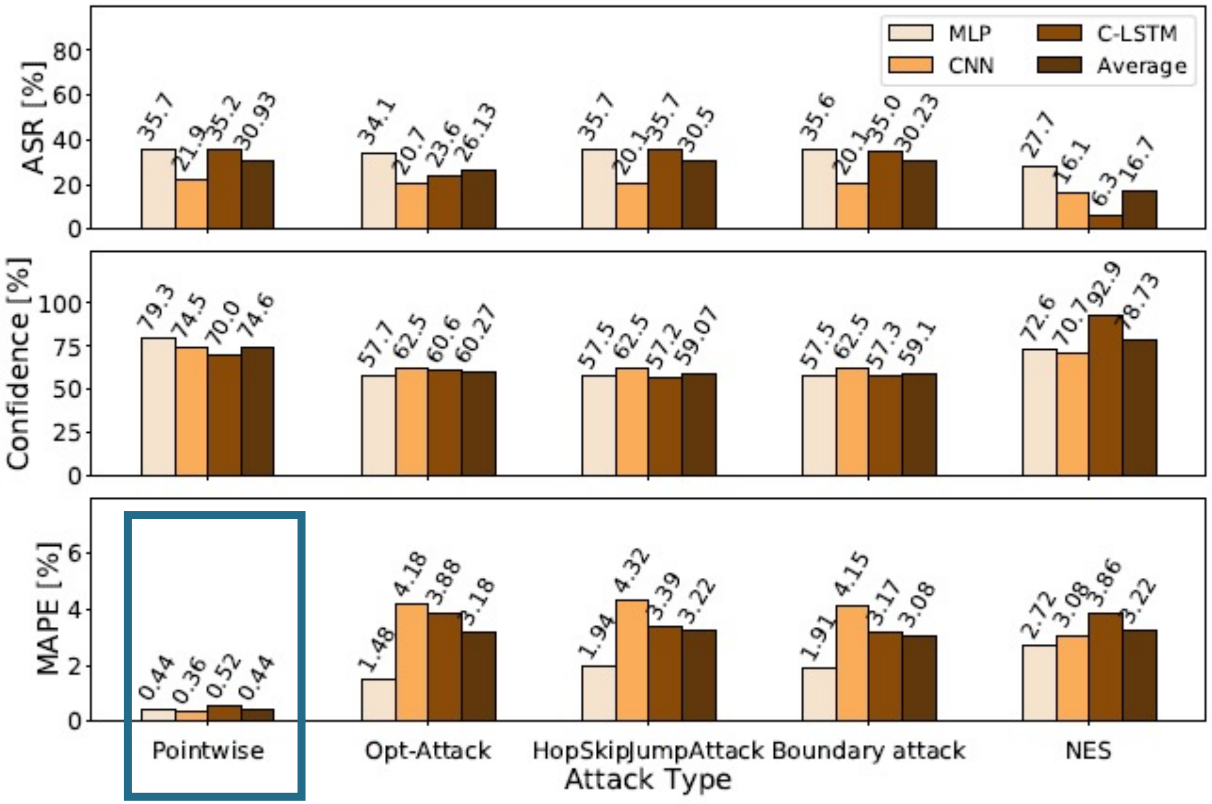
Adversarial attack performance



- Average attack success rates up to 35.7%

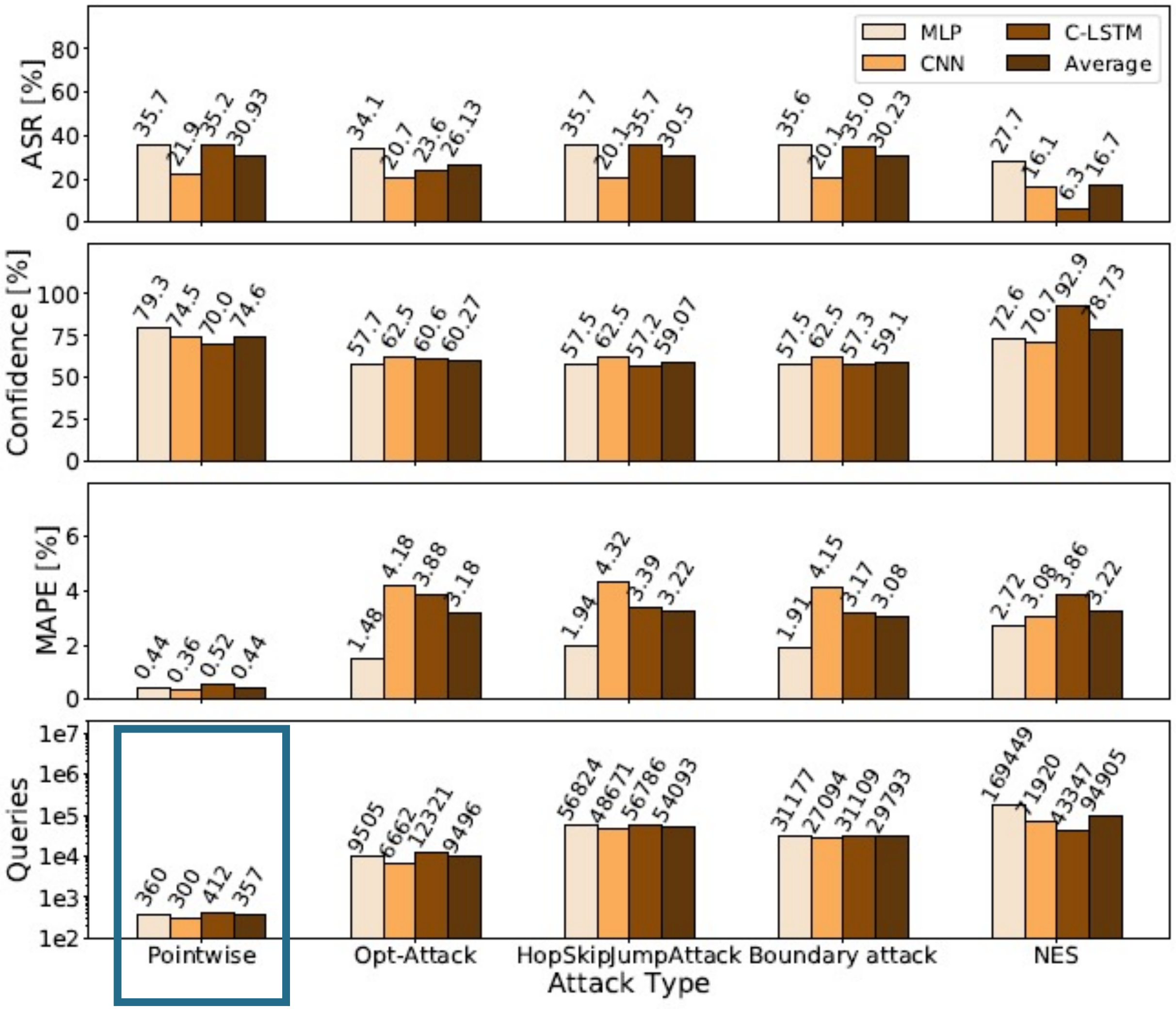
- **NIDS subverted while having up to 92.9% confidence in decision**

Adversarial attack performance



- Average attack success rates up to 35.7%
- NIDS subverted while having up to 92.9% confidence in decision
- **The scale of the perturbations is subtle**

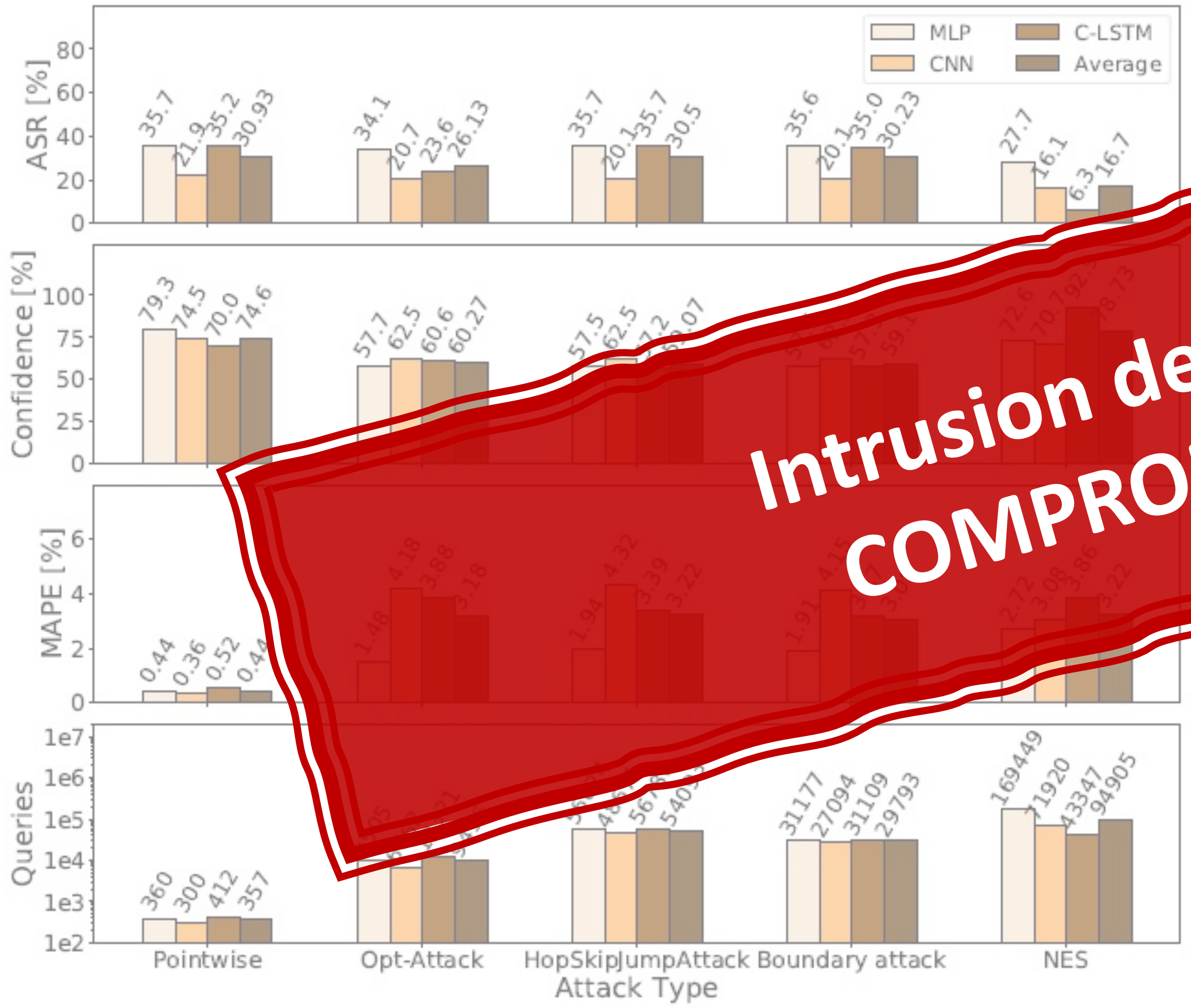
Adversarial attack performance



- Average attack success rates up to 35.7%
- NIDS subverted while having up to 92.9% confidence in decision
- The scale of the perturbations is subtle
- **As little as 300 queries needed to succeed**

P. Patras, *IoT Security on the Edge*

Adversarial attack performance

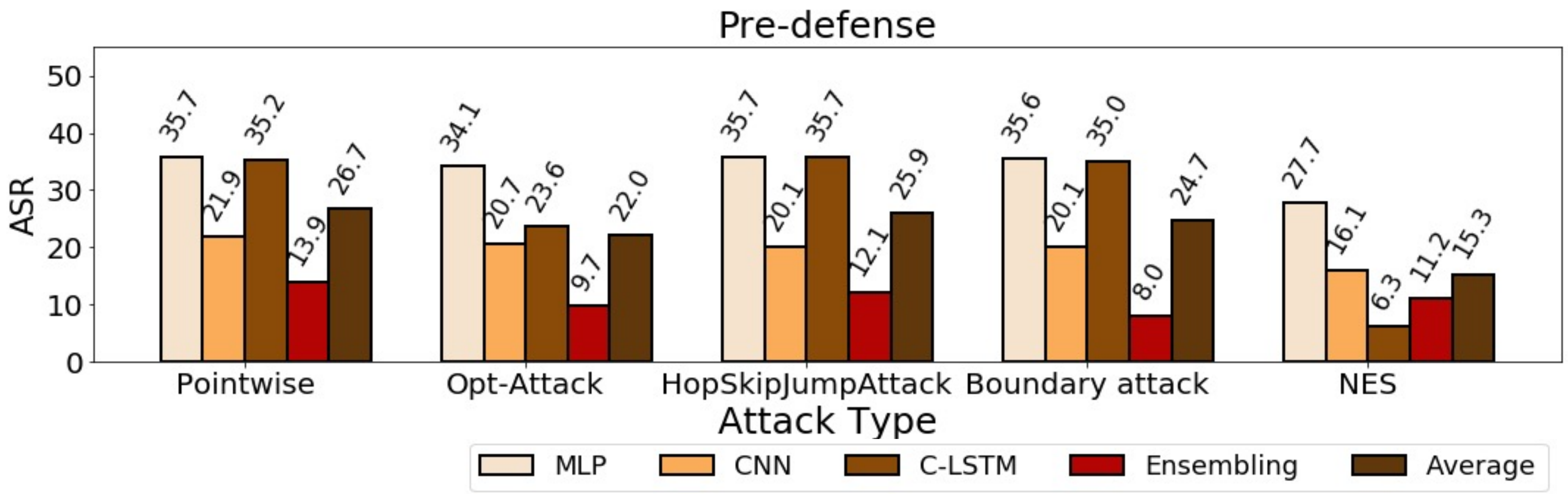


Intrusion detection COMPROMISED

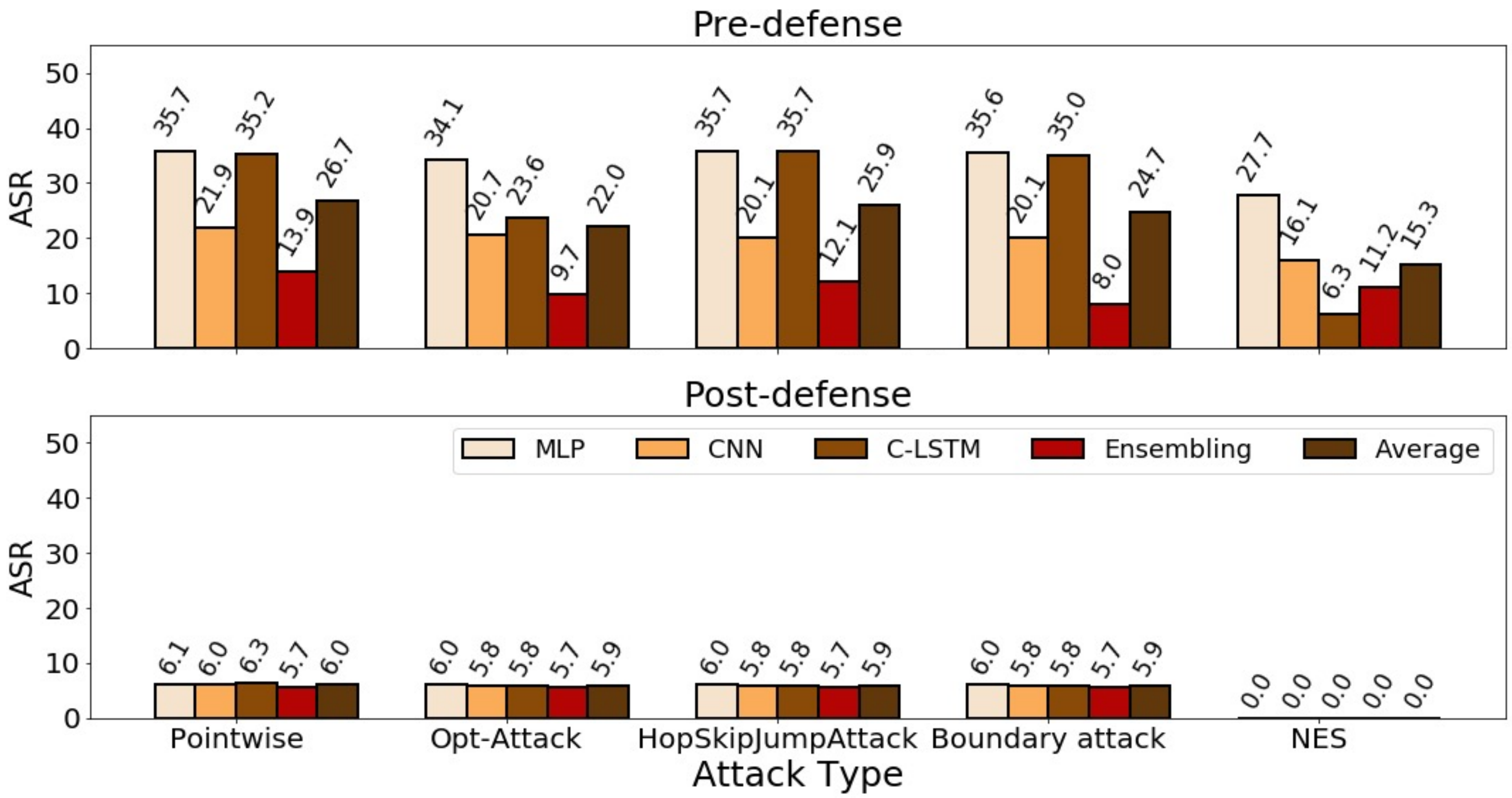
- Average attack success rates up to 35.7%
- NIDS subverted while having up to 92.9% confidence in decision
- The scale of the perturbations is subtle
- As little as 300 queries needed to succeed

P. Patras, *IoT Security on the Edge*

Performance after introducing defenses



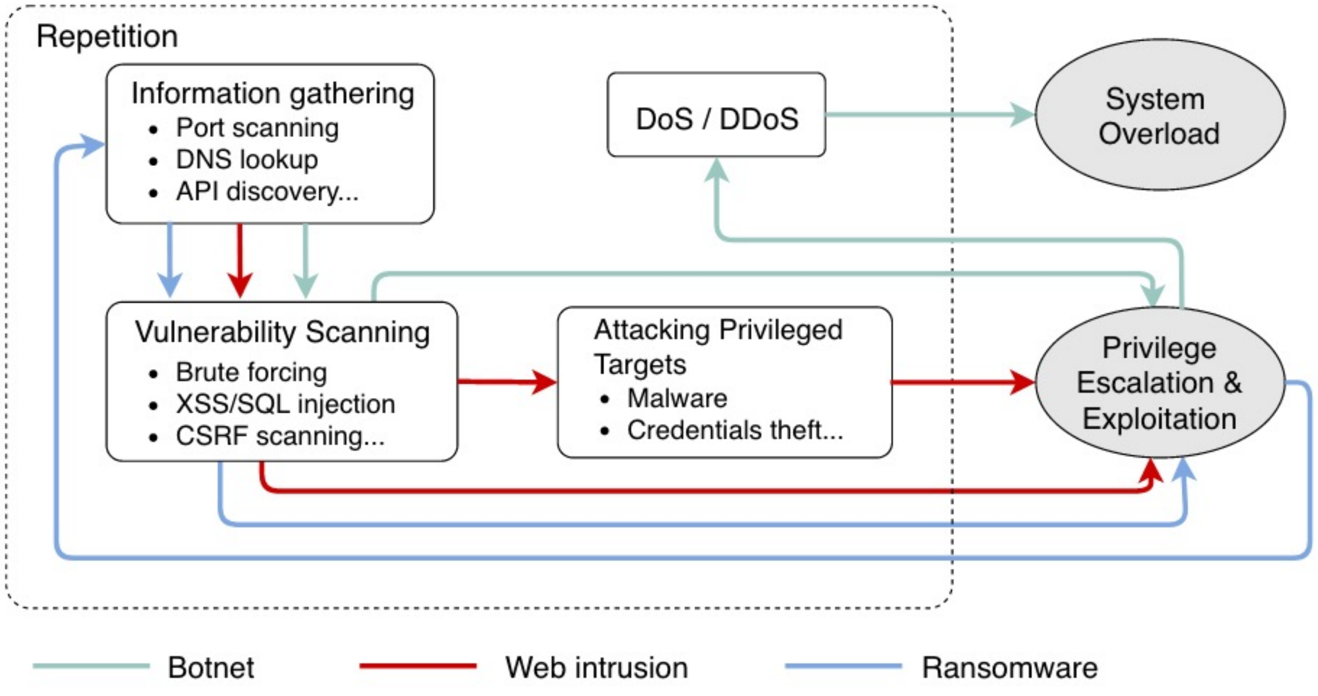
Performance after introducing defenses



ASR drops significantly after defenses applied

Can we build smarter lines of defense?

- What if you could exploit temporal ML models to detect threats before attacks proliferate?

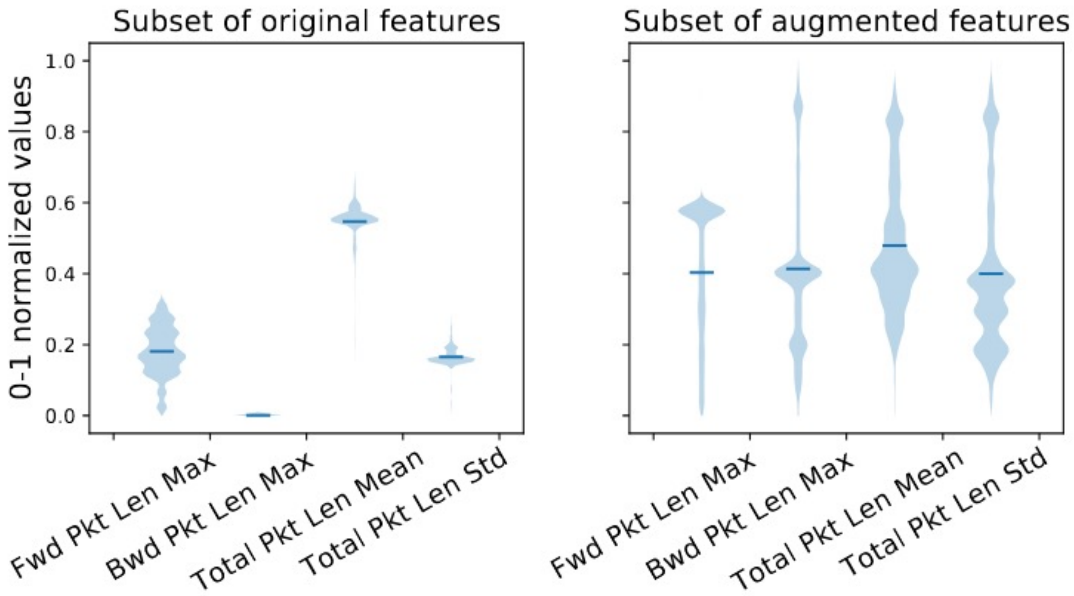


common stages shared by different large-scale cyber attacks

P. Patras, *IoT Security on the Edge*

Feature augmentation is key

- Training data largely collected in controlled environments
  → no accurate view of real-world network threats
- Models learn superficially and cannot generalise well



Bidirectional Asymmetric LSTM

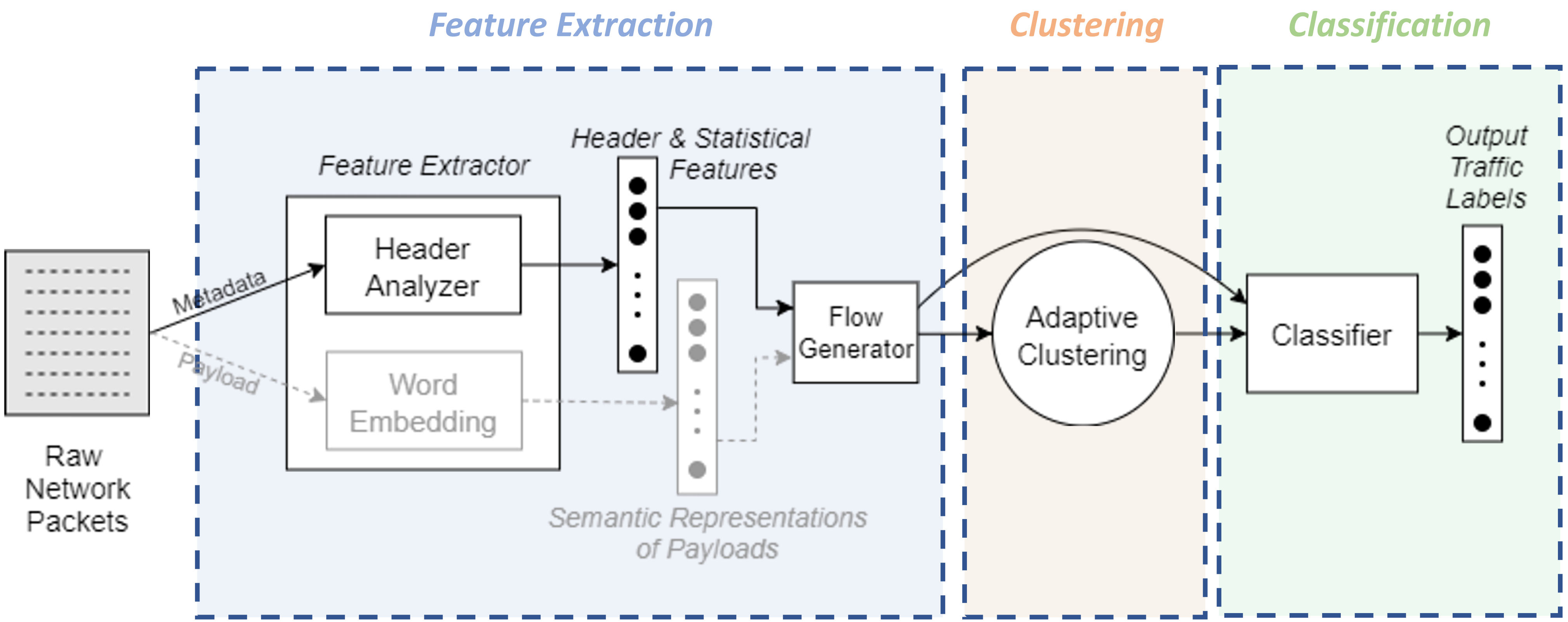
- Train two separate LSTM units, one for each processing direction
- Use future context to help the downstream classification task
- Different structures produce hidden states with different dimension (avoid redundancy)

Algorithm	CSE-CIC-IDS2018			CIC-IDS-2017 (X-eval)		
	precision	recall	F1	precision	recall	F1
RIPPER	0.9983	0.0981	0.1786	0.0873	0.0106	0.0190
Decision Tree	0.9989	0.9990	0.9990	0.5385	0.3717	0.4398
MLP	0.9989	0.9962	0.9976	0.6736	0.4631	0.5435
CNN	0.9947	0.9951	0.9949	0.7705	0.6344	0.6958
Autoencoder	0.7783	0.7500	0.7639	0.4362	0.4197	0.4278
OC-NN	0.9722	0.5310	0.6868	0.7844	0.5136	0.6208
Kitsune	0.6310	0.6081	0.6193	0.4086	0.3932	0.4007
DAGMM	0.8666	0.8253	0.8454	0.4159	0.3116	0.3576
Bi-LSTM	0.9990	0.9979	0.9985	0.7258	0.4209	0.5317
CNN-Bi-LSTM	**0.9996**	0.9982	0.9989	0.8813	0.3750	0.5261
Bi-ConvLSTM	0.9984	0.9971	0.9977	0.8721	**0.9693**	0.9178
Bi-ALSTM	0.9994	**0.9990**	**0.9992**	**0.9116**	0.9446	**0.9275**

- Bi-ALSTM generalizes remarkably well to previously unseen data
- Feature augmentation boosts performance of other models

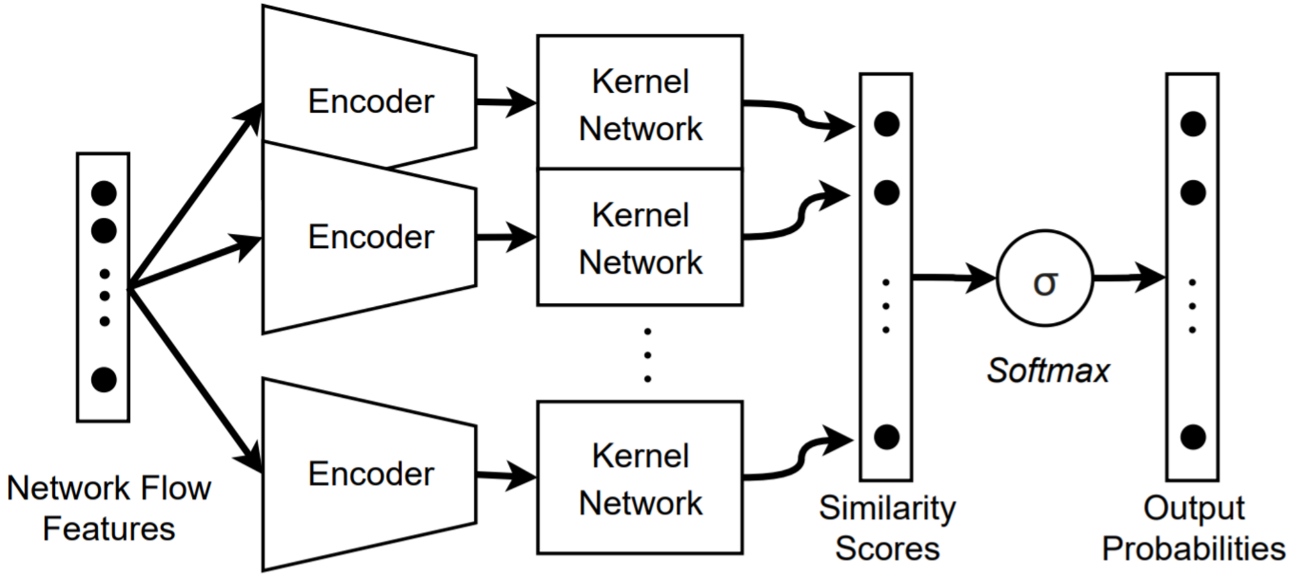
Can we do reliable NID at the edge?

ACID: Adaptive Clustering-based Intrusion Detection



A. Diallo and P. Patras, "Adaptive Clustering-based Malicious Traffic Classification at the Network Edge", IEEE INFOCOM 2021.

Adaptive Clustering network (AC-Net)



Key Advantages:

- Highly parallelizable
- Small computation/memory requirements
- Optimal separation of different classes
  also adaptive to complex and intertwined data structures
- Learns cluster centers on the fly

Performance

- 100% accuracy
- 0% false alarm rate
  (even 0.1% would be too high at current traffic speeds)
- 100% F1-score
- Inference time/sample:
  - 0.78 ms (without payload features)
  - 145ms (with payload features)
- Batch processing gives 100x speed-ups

Summary

Widespread technology broken. Can we change/amend standards?

Lots of work remains to be done to improve traffic classification robustness

Pioneering work on deep learning-based NIDS and defending against adversarial attacks

Hardware support essential for deploying ML at the edge for security

Additional research on traffic analysis and mobile security & privacy