Anomaly-based intrusion detection: challenges and possible strategies from unknowns to APT detection

Andrea Ceccarelli

with the contribution of:

A. Bondavalli, T. Puccetti, T. Zoppi, and BsC and MsC students from the University of Florence.

UNIVERSITÀ DEGLI STUDI FIRENZE
DIMAI
DIPARTIMENTO DI MATEMATICA E INFORMATICA "ULISSE DINI"

RCL
RESILIENT COMPUTING LAB

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

My introduction

Computer scientist with +15 years of experience in the **design and evaluation of dependable and secure systems**

- With case studies from railway, automotive, smart grid, industrial automation, software-intensive systems

RCL

RESILIENT COMPUTING LAB

https://rcl.unifi.it

Not a «machine learning guy»

- Enabling technology to reach our goal

Presentation Outline

1. Context: why and how anomaly-based intrusion detection

2. Which classifier
   - The role of DNNs
   - Detection of unknown attacks (zero-days)
   - Take advantage of many: stacking

3. A forgotten measure: attack latency

4. What's next: defend against Advanced Persistent Threats

Presentation Outline

1. Context: why and how anomaly-based intrusion detection

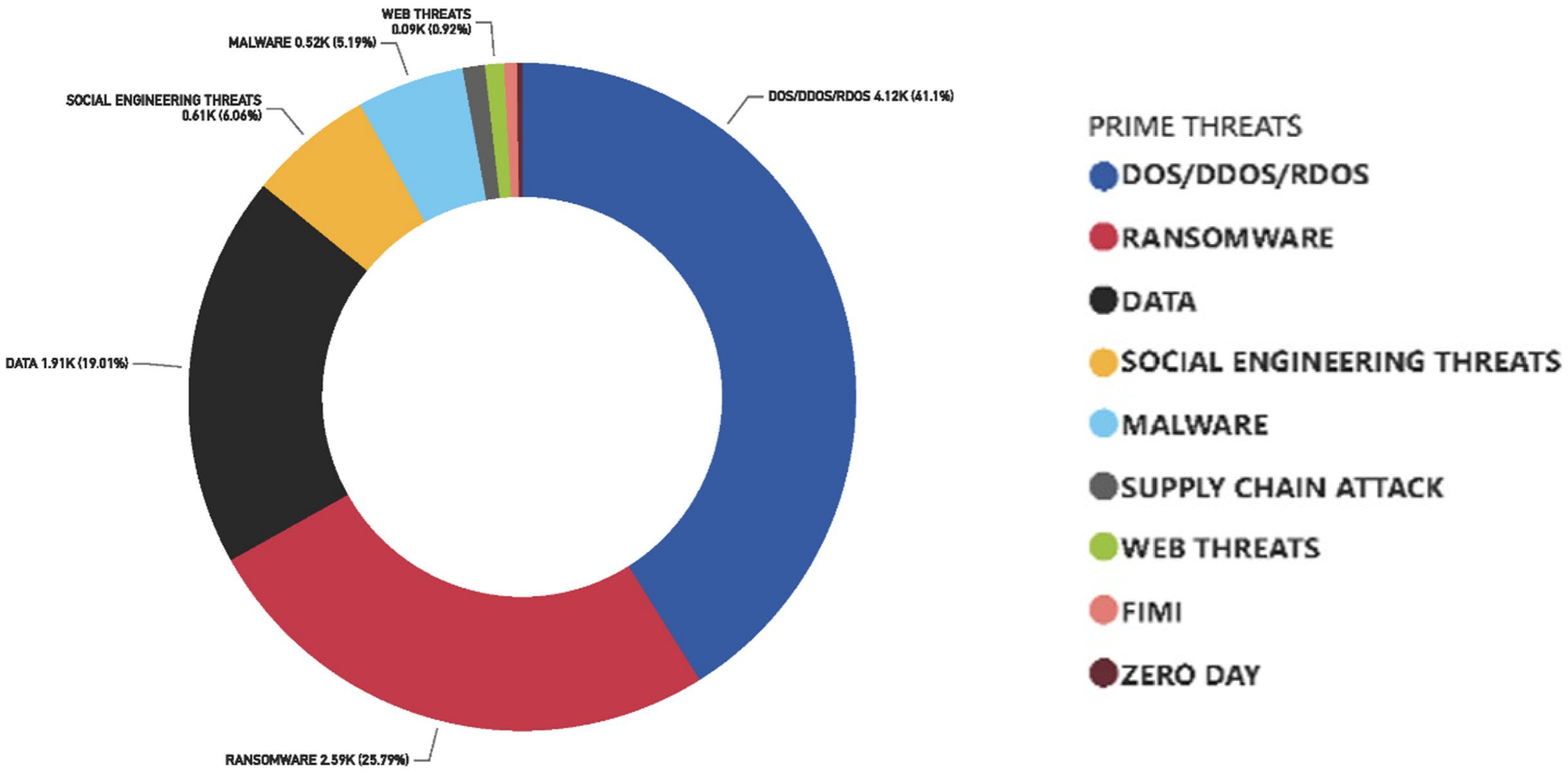
2. Which classifier
   - The role of DNNs
   - Detection of unknown attacks (zero-days)
   - Take advantage of many: stacking

3. A forgotten measure: attack latency

4. What's next: defend against Advanced Persistent Threats

ENISA's Threat Landscape - analyzed incidents by threat type

Violations to confidentiality, availability, integrity



https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024

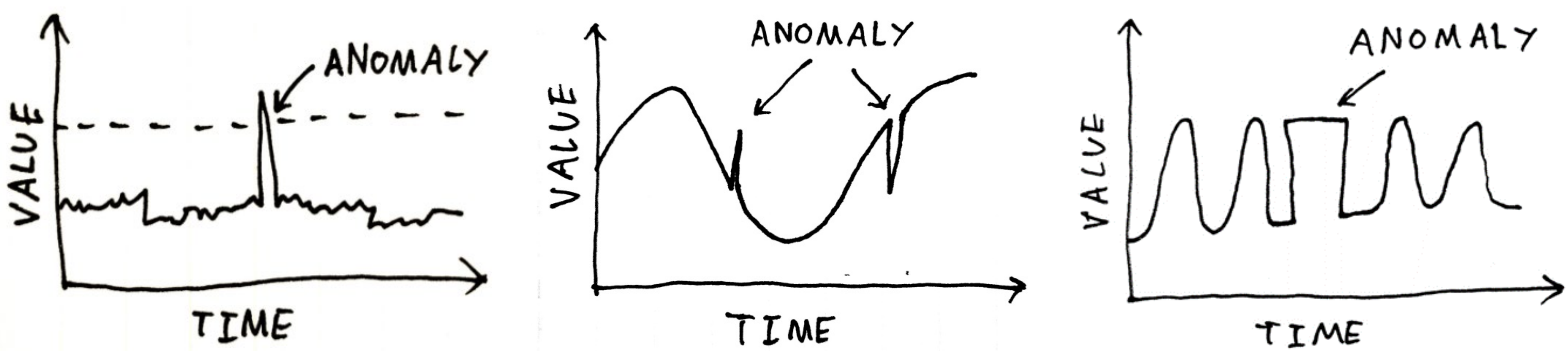
How to defend

Means to realize intrusion detections:

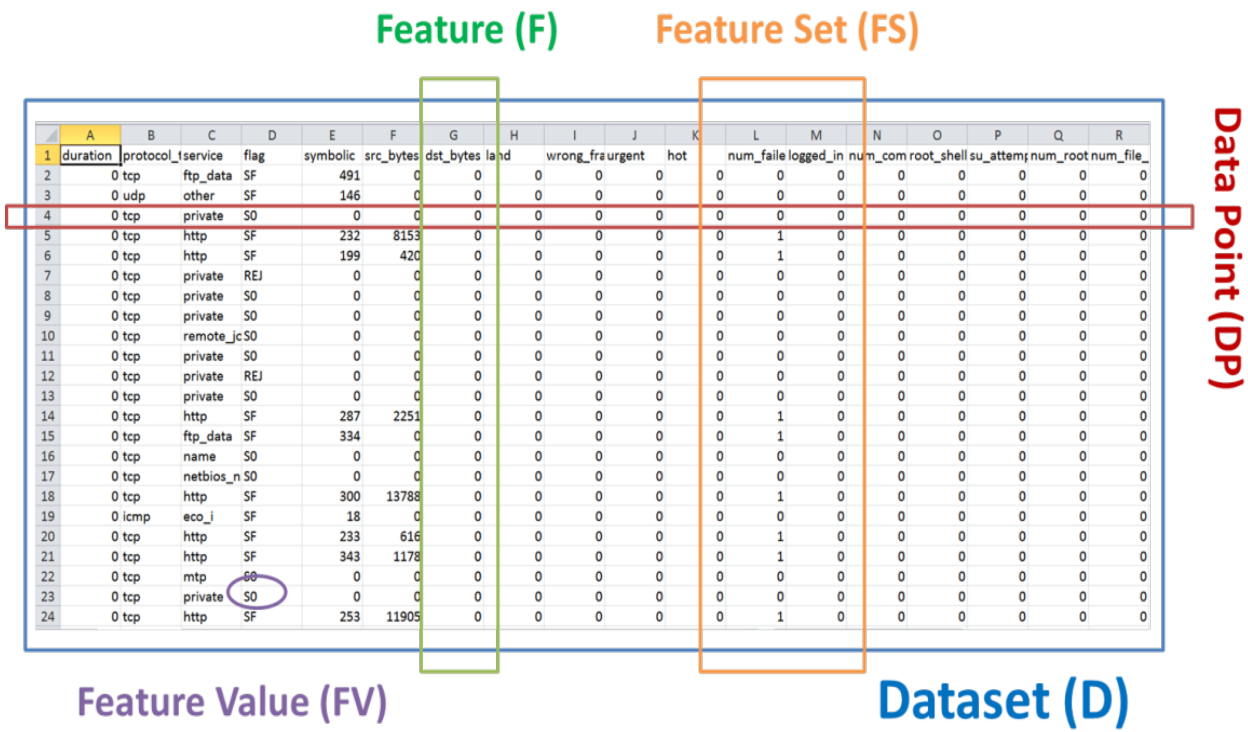
Rule-based, Invariant-Based, Signature-based

our focus!

Anomaly-based (under the underlying assumption that attacks have a visible effect on monitored system indicators)



It is just binary classification on tabular data



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	duration	protocol_t	service	flag	symbolic	src_bytes	dst_bytes	land	wrong_fra	urgent	hot	num_faile	logged_in	num_com	root_shell	su_attemp	num_root	num_file_
2	0	tcp	ftp_data	SF	491	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	udp	other	SF	146	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	tcp	http	SF	232	8153	0	0	0	0	0	1	0	0	0	0	0	0
6	0	tcp	http	SF	199	420	0	0	0	0	0	1	0	0	0	0	0	0
7	0	tcp	private	REJ	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	tcp	remote_jc	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	tcp	private	REJ	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	tcp	http	SF	287	2251	0	0	0	0	0	1	0	0	0	0	0	0
15	0	tcp	ftp_data	SF	334	0	0	0	0	0	0	1	0	0	0	0	0	0
16	0	tcp	name	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	tcp	netbios_n	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	tcp	http	SF	300	13788	0	0	0	0	0	1	0	0	0	0	0	0
19	0	icmp	eco_i	SF	18	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	tcp	http	SF	233	616	0	0	0	0	0	1	0	0	0	0	0	0
21	0	tcp	http	SF	343	1178	0	0	0	0	0	1	0	0	0	0	0	0
22	0	tcp	mtp	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	tcp	http	SF	253	11905	0	0	0	0	0	1	0	0	0	0	0	0

Usually needs shuffling!
(loss of context?)

pcap **session summaries** **syscall traces**

system indicators **network indicators**

CTU-13 NSL-KDD ISCX12 UNSW-NB15 IoT-IDS

CICIDS17 AndMal17 SDN20

CICIDS18 Netflow-IDS

Need ad-hoc solutions?

	Malware	Web Attack	Web Application	Spam / Phishing	(D)Dos	BotNet	Data Breaches
NSL-KDD	u2r		r2l		DoS		Probe
CTU-13						BotNet	
ISCX12		BruteForce			DoS, DDoS		Infiltration
UNSW-NB15	Worms	Fuzzers	Backdoor, Exploits, Shellcode		DoS		Analysis, Reconnaissance
UGR16				Blacklist, Spam	DoS	BotNet	Scan
NGIDS-DS	Malware, Worms		Backdoor, Exploits, Shellcode		DoS		Reconnaissance
Netflow-IDS				Mailbomb	Neptune, Portsweep		
AndMal17	Ransomware, Scareware			SMS, Adware			
CIDDS-001		BruteForce			DoS		PortScan, PingScan
CICIDS17		BruteForce			DoS (Slowloris, Goldeneye)		PortScan
CICIDS18		BruteForce (FTP, SSH)			DoS, DDoS	Bot	Infiltration
SDN20		BruteForce	Exploits		DoS, DDoS		Probe

different features **different systems** **same attack, different visible effects**

Catillo, Marta, et al. "Transferability of machine learning models learned from public intrusion detection datasets: the CICIDS2017 case study." Software Quality Journal 30.4 (2022): 955-981.

T. Zoppi, et al. "Towards a general model for intrusion detection: An exploratory study." *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Cham: Springer Nature Switzerland, 2022.

Presentation Outline

1. Context: why and how anomaly-based intrusion detection
2. Which classifier
   – The role of DNNs
   – Detection of unknown attacks (zero-days)
   – Take advantage of many: stacking
3. A forgotten measure: attack latency
4. What's next: defend against Advanced Persistent Threats

Let's start training and testing!

Supervised: labels are used when training

XGBoost, Random Forests, LDA, Knn, ExtraTrees, …

Unsupervised: no labels during training

Isolation Forest, FastAbod, K-means, ODIN, …

	Known Events	Unknown Events
Supervised	Very Good!	Potentially Bad
Unsupervised	Average	

attacks!

Which supervised?

	Known Events	Unknown Events
Supervised	Very Good!	Potentially Bad
Unsupervised	Average	

attacks

Nowadays DNNs are very popular as they work well in many applications

However, efficacy unclear for tabular data

Shwartz-Ziv, Ravid, and Amitai Armon. "Tabular data: Deep learning is not all you need." Information Fusion 81 (2022): 84-90.

Ye, Han-Jia, et al. "A closer look at deep learning on tabular data." *arXiv preprint arXiv:2407.00956* (2024).

In case of IDS?

T. Zoppi, et al. "Anomaly-based error and intrusion detection in tabular data: no DNN outperforms tree-based classifiers." Future Generation Computer Systems 160 (2024): 951-965.

Which supervised?

	Known Events	attacks	Unknown Events
Supervised	Very Good!		Potentially Bad
Unsupervised	Average		

23 datasets, attacks known at training time

DNN-based supervised algorithms FastAI, TabNet, NODE, GATE, …

Including image-based DNNs exploiting DeepInsight

Tree-based classifiers *Random Forests*, *eXtreme Gradient Boosting (XGBoost)* or *Extra Trees* outperform DNNs

– also easier to fine-tune, and understand

– less time and resources to learn their model

▸ True independently on the dimension of the training set

With unknowns?

	Known Events	Unknown Events
Supervised	Very Good	Potentially Bad
Unsupervised	Average	

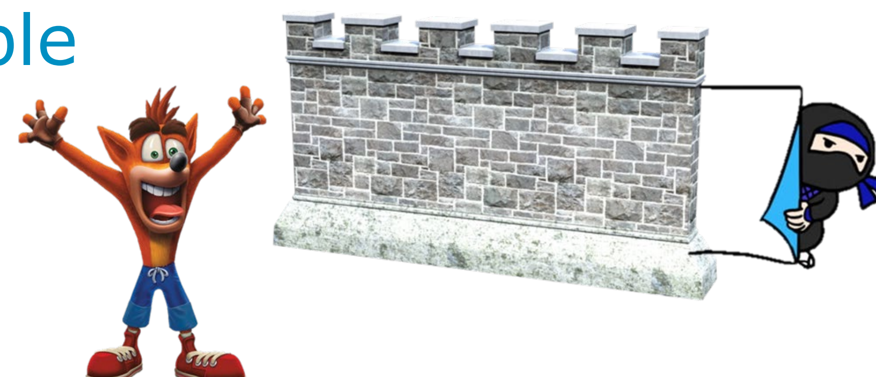
ttacks!

Research and Practice found ways to defend against specific attacks

- Mostly rule, signature-based or supervised (tree-based) learning

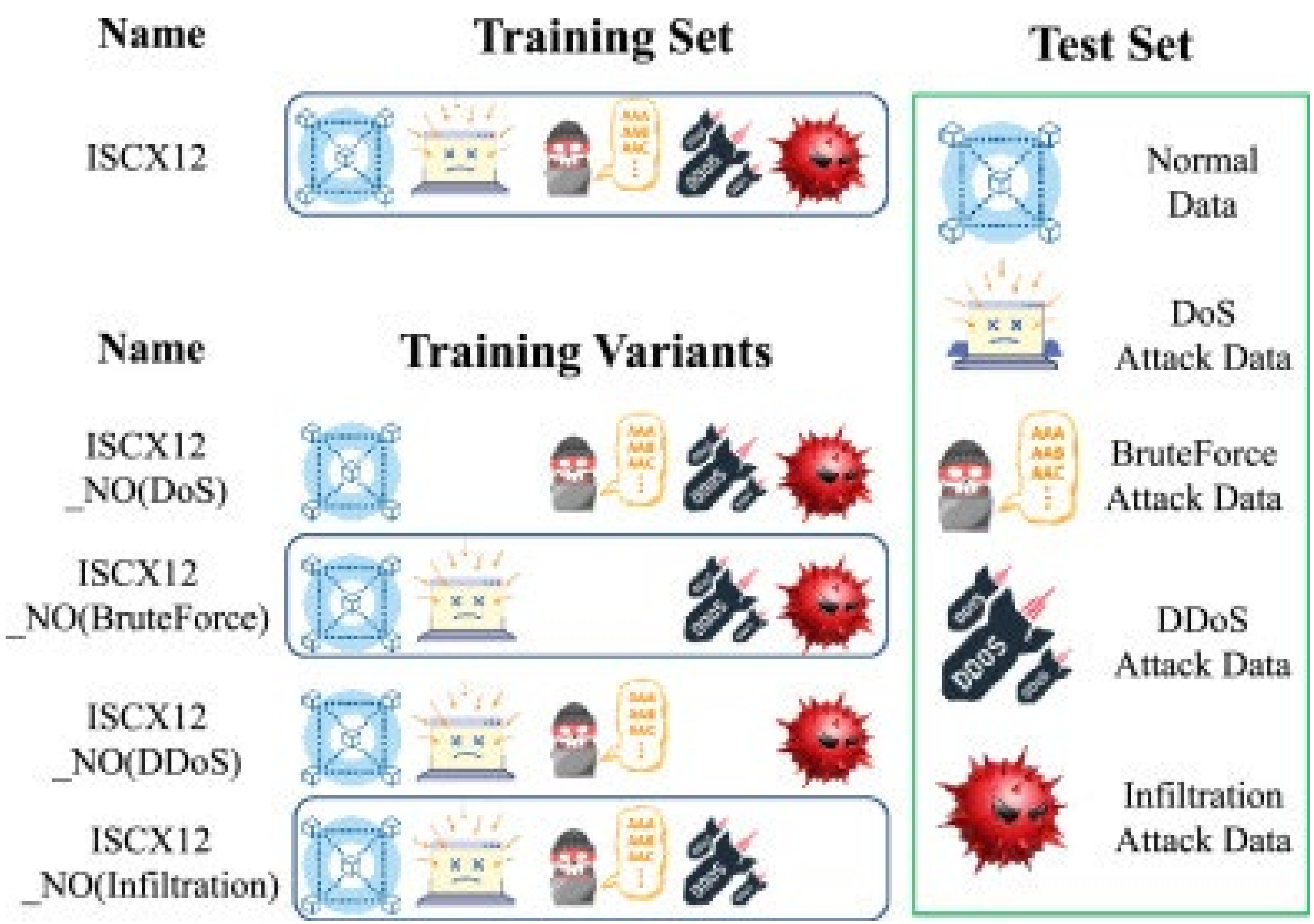
But what about with zero days, variants, … ?

- No rule / signature available
- Anomaly detectors much less efficient



How to test?

	Known Events	Unknown Events
Supervised	Very Good	Potentially Bad
Unsupervised	Average	



Zoppi, Tommaso, et al. "Which algorithm can detect unknown attacks? Comparison of supervised, unsupervised and meta-learning algorithms for intrusion detection." *Computers & Security* 127 (2023): 103107.

Datasets Variants

	Known Events	Unknown Events
Supervised	Very Good	Potentially Bad
Unsupervised	Average	

Name	Year	# Data Points	Features		Attacks		# Variants
			Ord.	Cat.	#	%	
ADFANet	2015	132 002	5	6	3	11.3	3
AndMal17	2017	100 000	77	5	4	15.5	4
CICIDS17	2017	500 000	77	5	5	79.7	5
CICIDS18	2018	200 000	77	5	8	26.2	8
CIDDS	2015	400 000	5	7	4	14.4	4
IoT-IDS	2019	210 425	8	1	8	42.3	8
ISCX12	2012	600 000	4	10	4	43.5	4
NSLKDD	2009	148 516	37	5	4	40.7	4
SDN20	2020	205 167	63	5	5	66.6	5
UGR16	2016	207 256	4	6	5	3.3	5
UNSW-NB15	2015	165 461	38	6	8	6.5	8

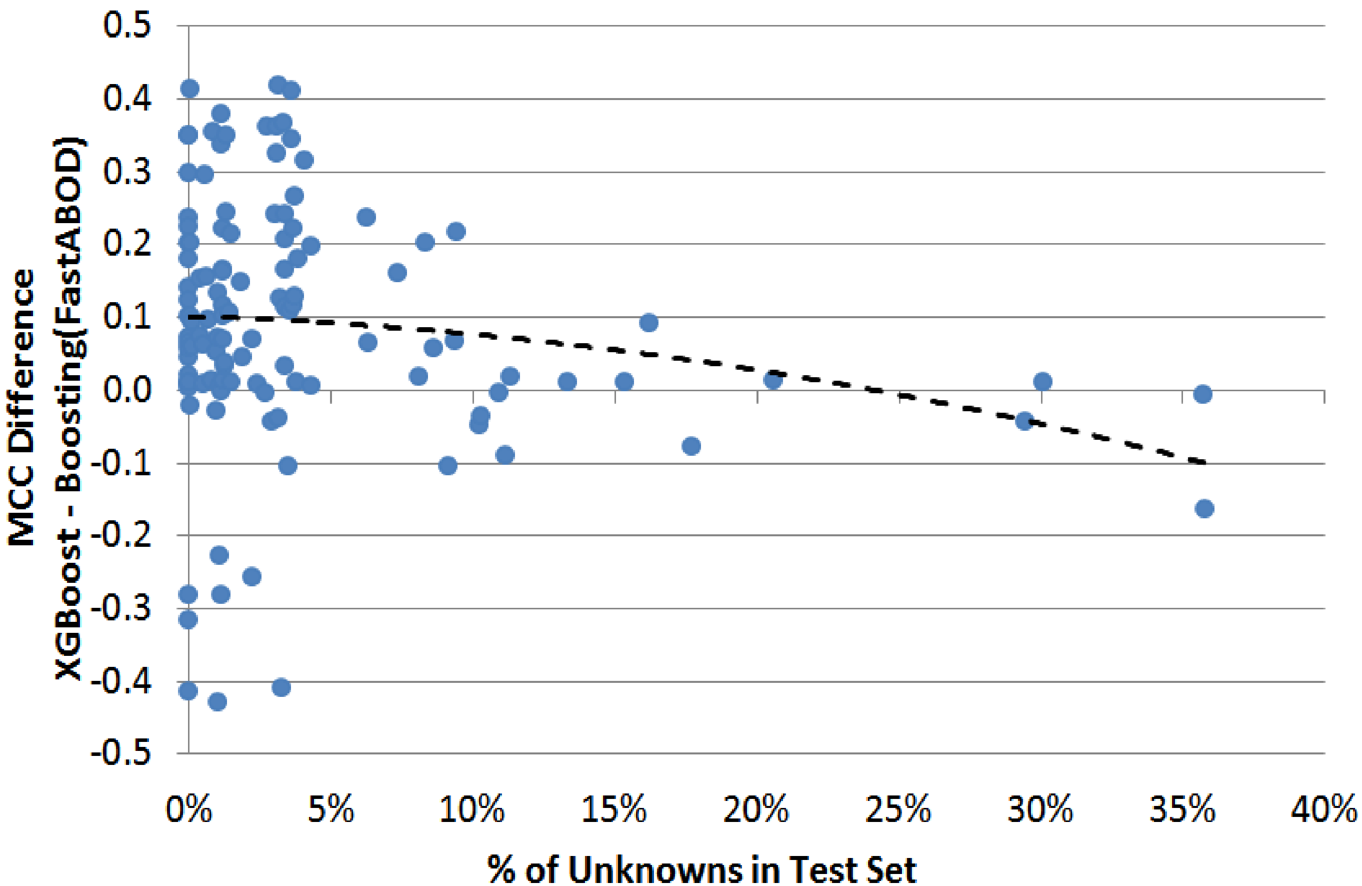
Some of the attack datasets we used

- the more attacks a dataset contains, the more variants

… and all the data!

	Known Events	Unknown Events
Supervised	Very Good	Potentially Bad
Unsupervised	Average	

Differences between the best supervised and unsupervised algorithm, when varying the number of unknowns



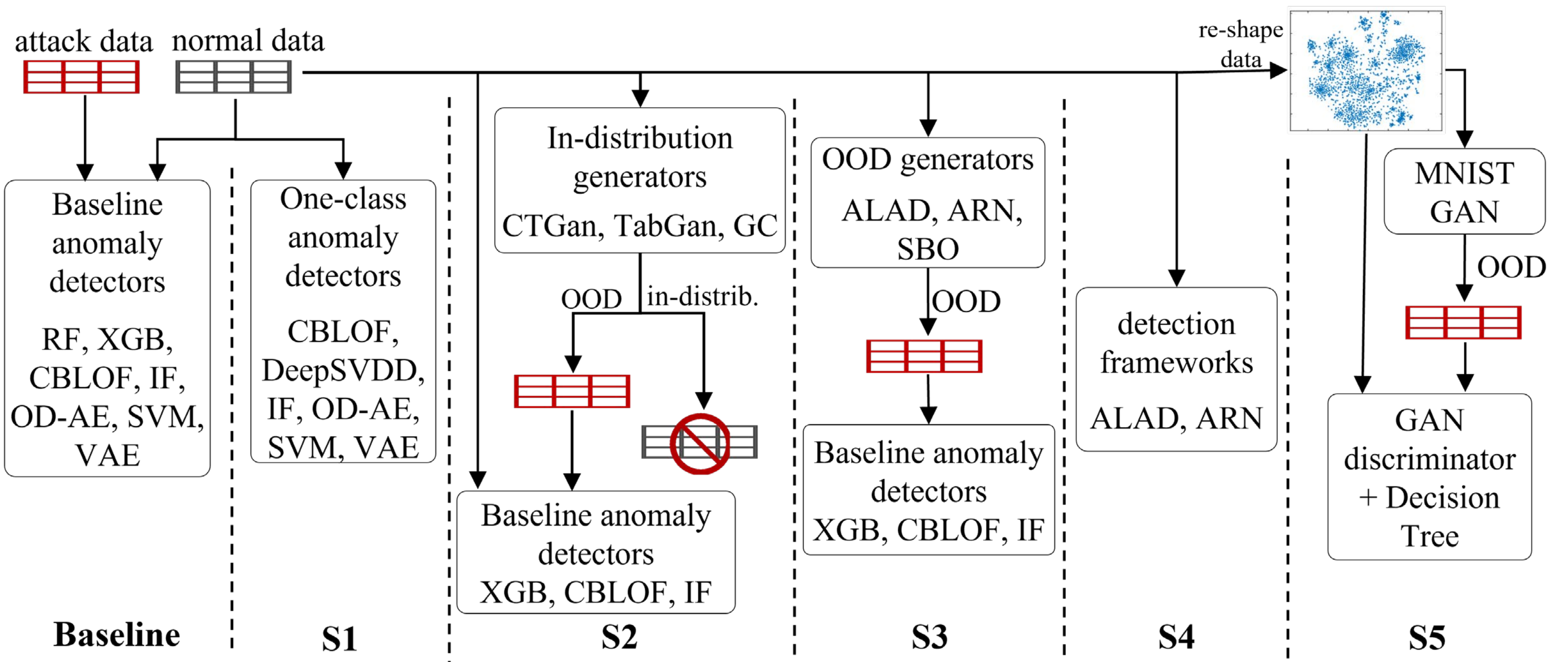
If zero knowledge?

	Known Events	Unknown Events
Supervised	Very Good	Potentially Bad
Unsupervised	Average	

ttacks!

Difficult to obtain good attack data

time-consuming, expensive, incomplete, outdated, etc.



But no alternatives– aside when few easy features

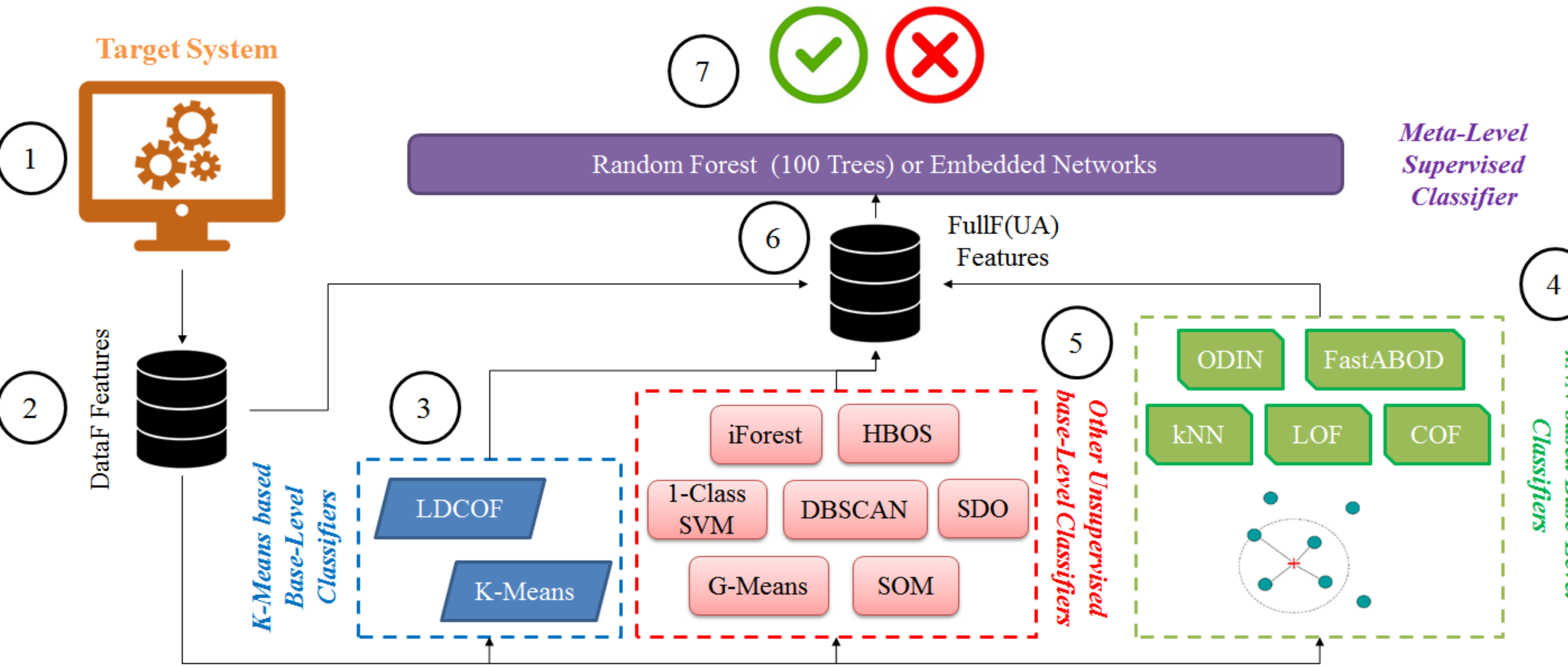
A. Ceccarelli, and T. Zoppi. "Intrusion detection without attack knowledge: generating out-of-distribution tabular data." ISSRE 2023

Ensembles: take the best from both!

attacks!

	Known Events	Unknown Events
Supervised	Very Good!	Potentially Bad
Unsupervised	Average	

Boosting, Bagging, Stacking!



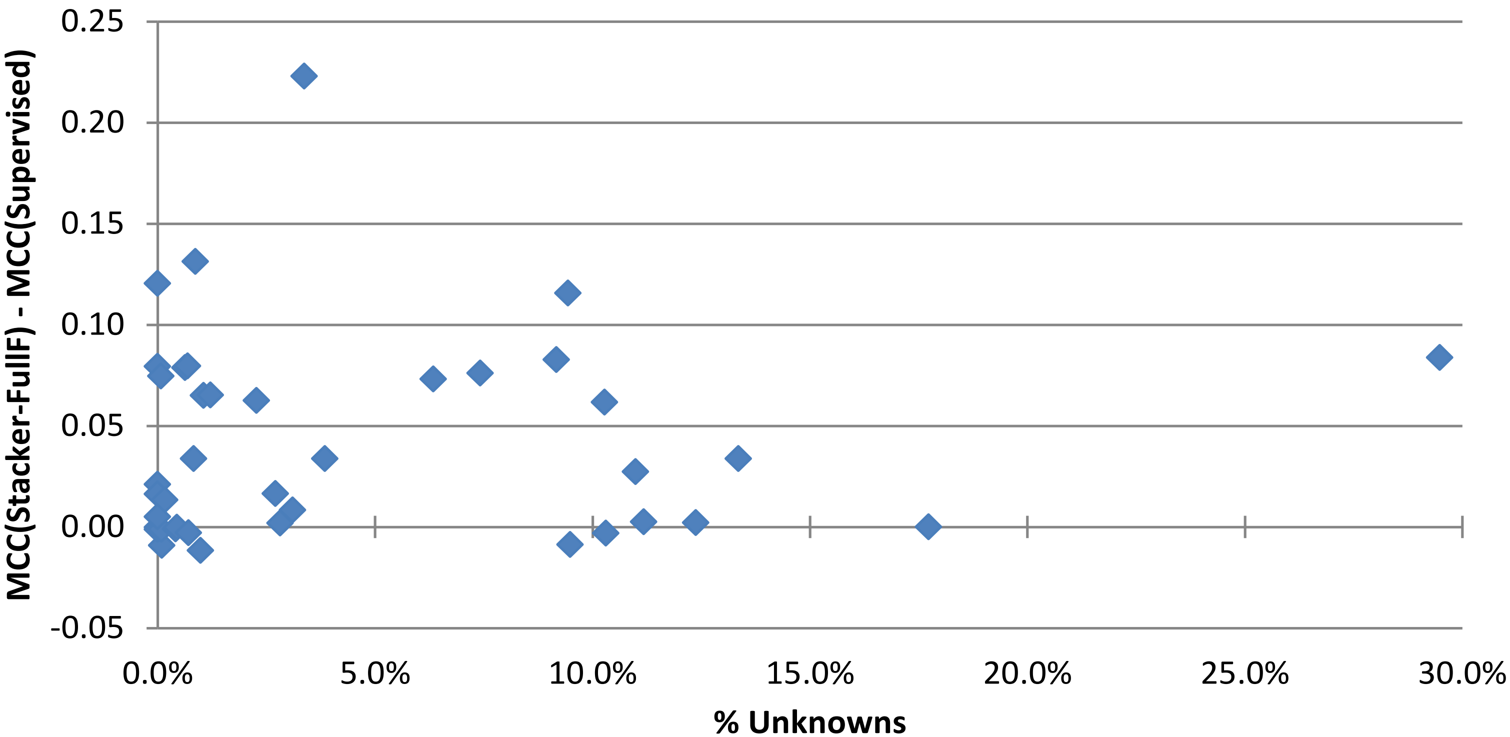
Zoppi, T., Ceccarelli, A. (2021) "Prepare for trouble and make it double! Supervised–Unsupervised stacking for anomaly-based intrusion detection." *Journal of Network and Computer Applications* 189: 103106.

Evaluation of the Stacker

	Known Events	attacks!	Unknown Events
Supervised	Very Good!		Potentially Bad
Unsupervised	Average		

Comparison between MCC Stacker vs **supervised**

Each dataset, we take the best supervised algorithm



Presentation Outline

1. Context: why and how anomaly-based intrusion detection

2. Which classifier
   – The role of DNNs
   – Detection of unknown attacks (zero-days)
   – Take advantage of many: stacking

3. A forgotten measure: attack latency

4. What’s next: defend against Advanced Persistent Threats

Metrics that makes us happy!

anomaly detection and tabular data in top dependability and security venues

Paper	Venue	Metrics
Jha et al. 2022	DSN	P, R, F1, Lead Detection Time.
Wang et al. 2022	DSN	P, R, F1
Dayaratne et al. 2022	DSN	P, R,F1, FPR
Alharthi et al. 2021	DSN	P, R, F1, MCC
Yuan et al. 2021	DSN	P, R, TPR, FPR
Xu et al. 2021	DSN	P, R, F1
Zhao et al. 2019	DSN	A
Wang et al. 2022	ISSRE	P, R, F1
Zhang et al. 2021	ISSRE	P, R, F1
Jia et al. 2021	ISSRE	P, R
Zhang et al. 2021	ISSRE	P, R, F1,ROC
Alsaheel et al. 2021	USENIX	P, R F1, ROC
Chen et al. 2021	USENIX	R, avg. time
Downing et al. 2021	USENIX	P, R, FPR, ROC
Izhikevich et al. 2021	USENIX	A, proc. time
Fu et al. 2021	USENIX	P, R, FPR
Tang et al. 2021	USENIX	TPR, FPR

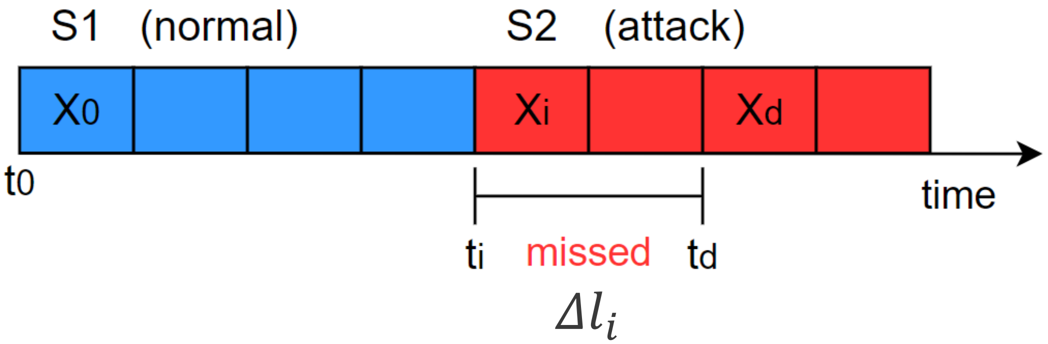
What is usually studied are anomalies represented by individual data points, observed in datasets composed by hours of normal concatenated with hours of attacks.

Are we forgetting attack latency?

How long was the attacker into the system before being detected?

Or: given a complex attack, how long did it take to detect it?

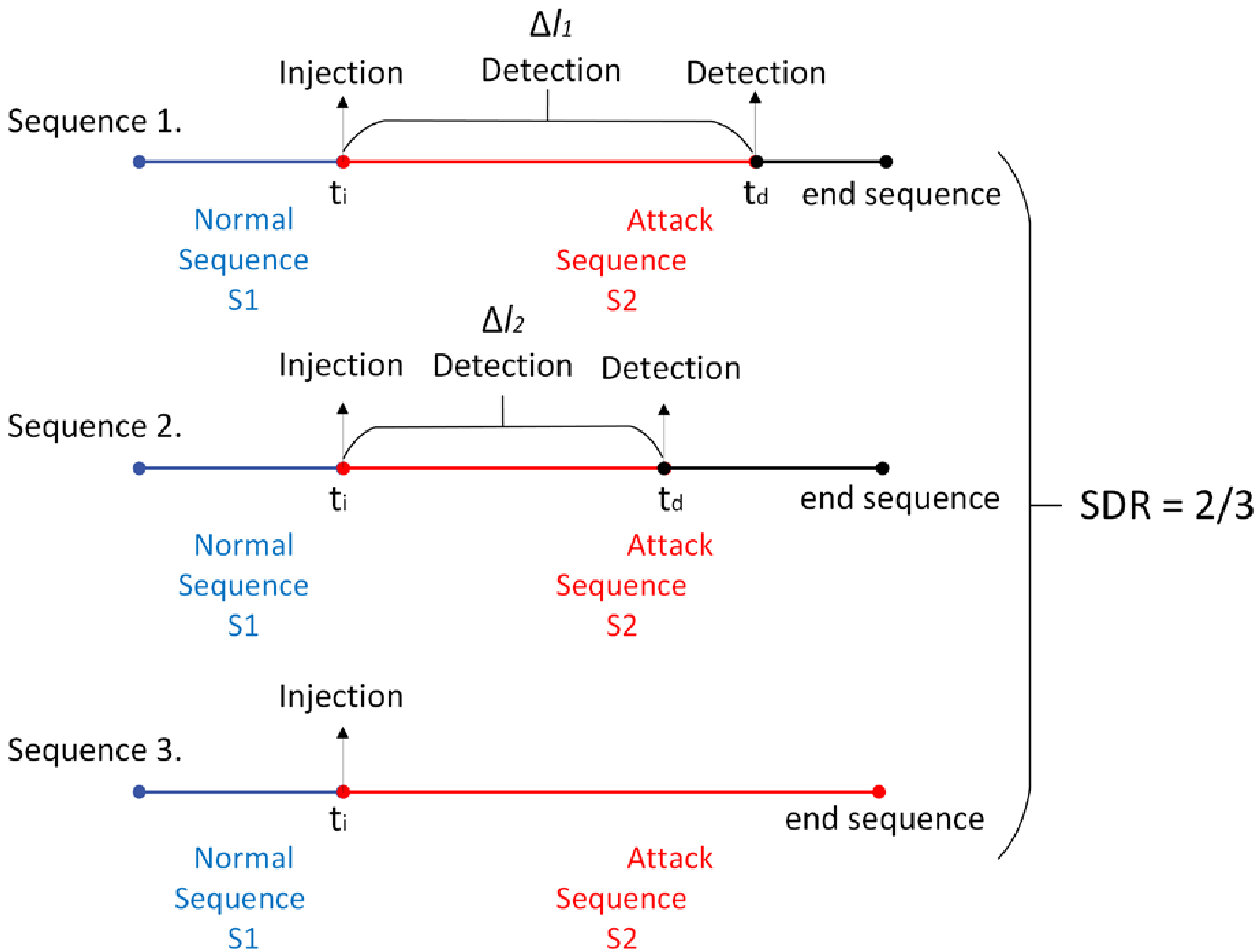
- **Average Latency** = $\boldsymbol{\Delta L} = \frac{\sum_{i=0}^{N} \Delta l_i}{N}$
- **Sequence Detection Rate SDR** (as there is the case in which x_d never occur)



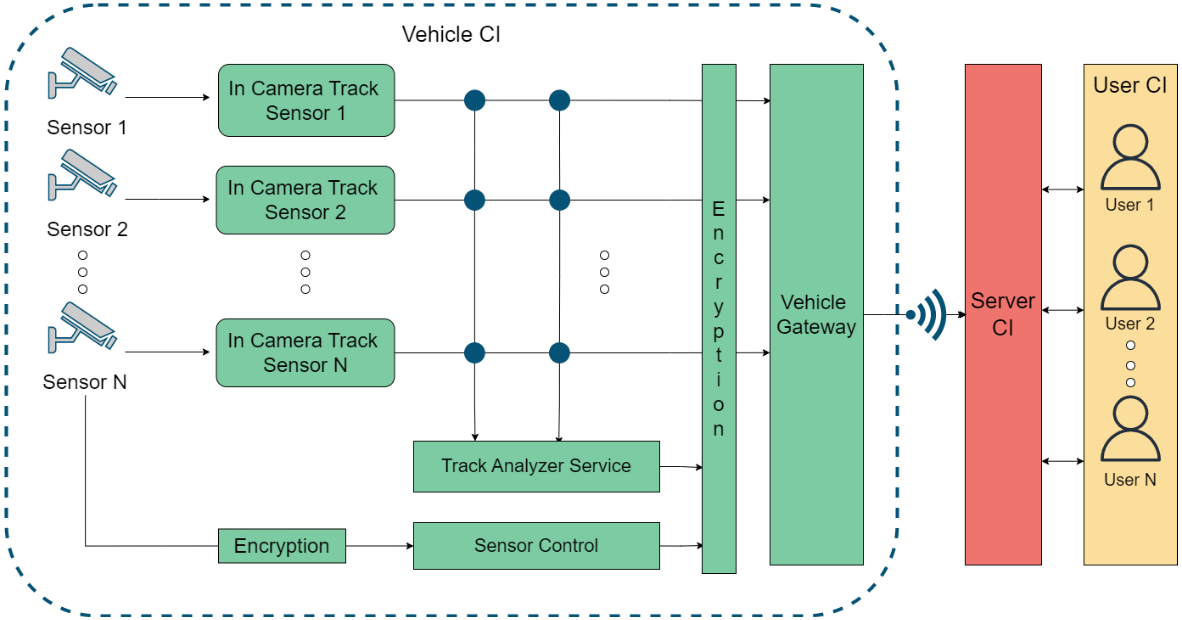
Tommaso Puccetti and Andrea Ceccarelli , Detection Latencies of Anomaly Detectors: An Overlooked Perspective?, *ISSRE 2024*

Puccetti, T., Nardi, S., Cinquilli, C., Zoppi, T., & Ceccarelli, A. (2024). ROSPaCe: Intrusion Detection Dataset for a ROS2-Based Cyber-Physical System and IoT Networks. *Scientific Data*, *11*(1), 481.

A bit more on the SDR

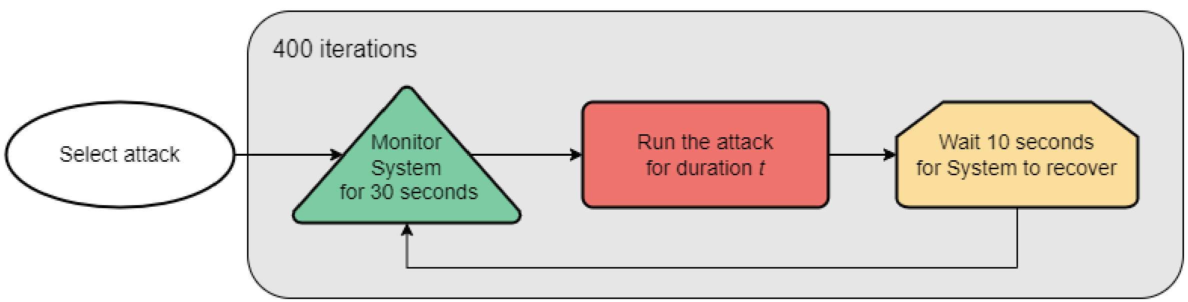


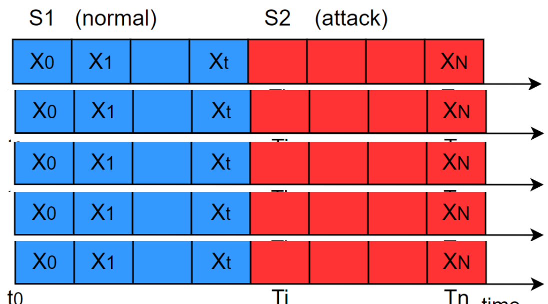
ROSPaCe data collection procedure



6 different attacks:

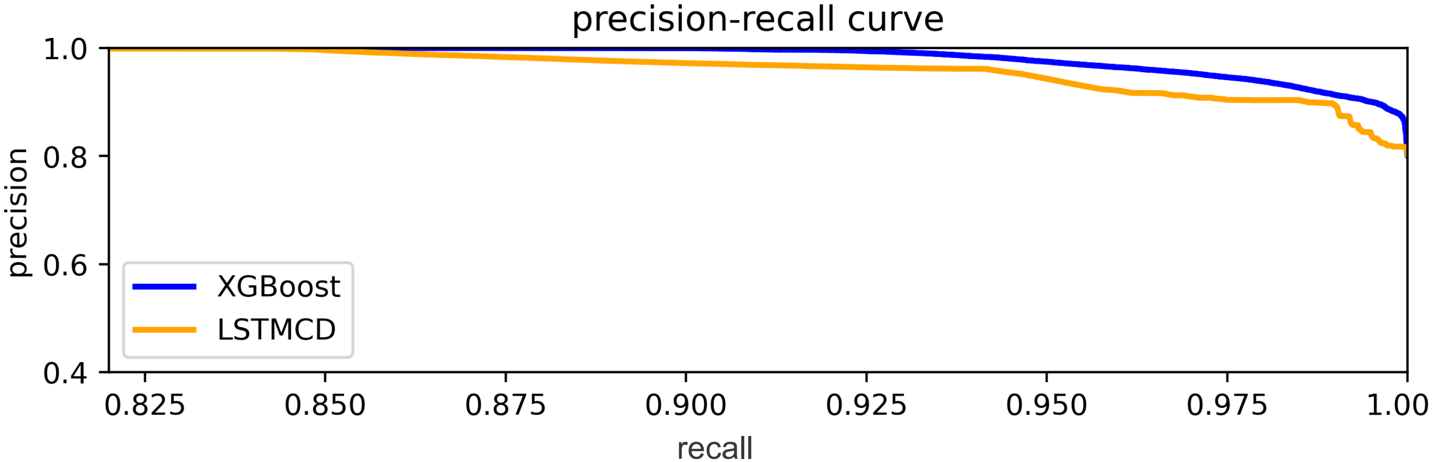
- 2 discovery attacks
- 4 DoS attacks

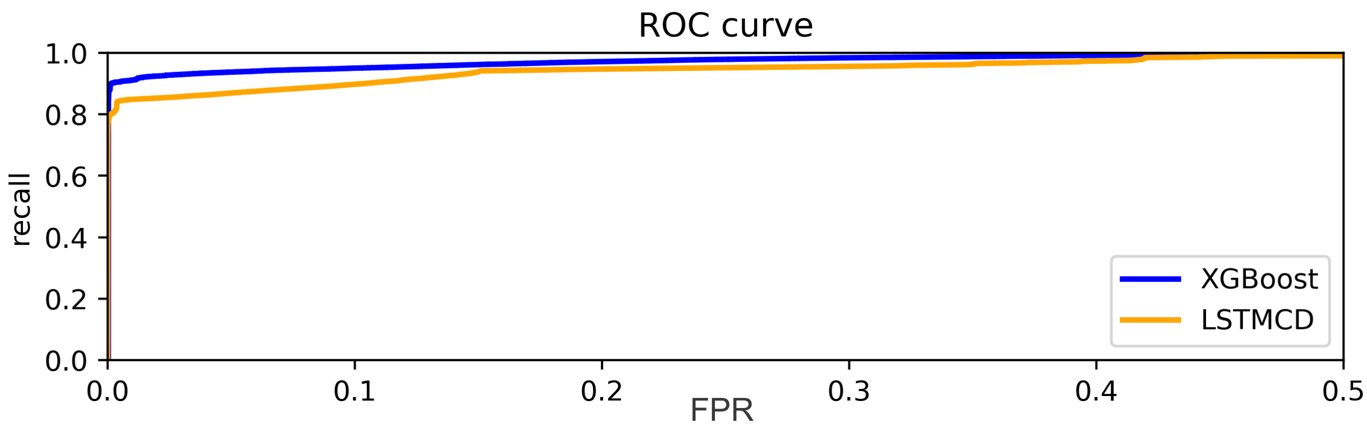




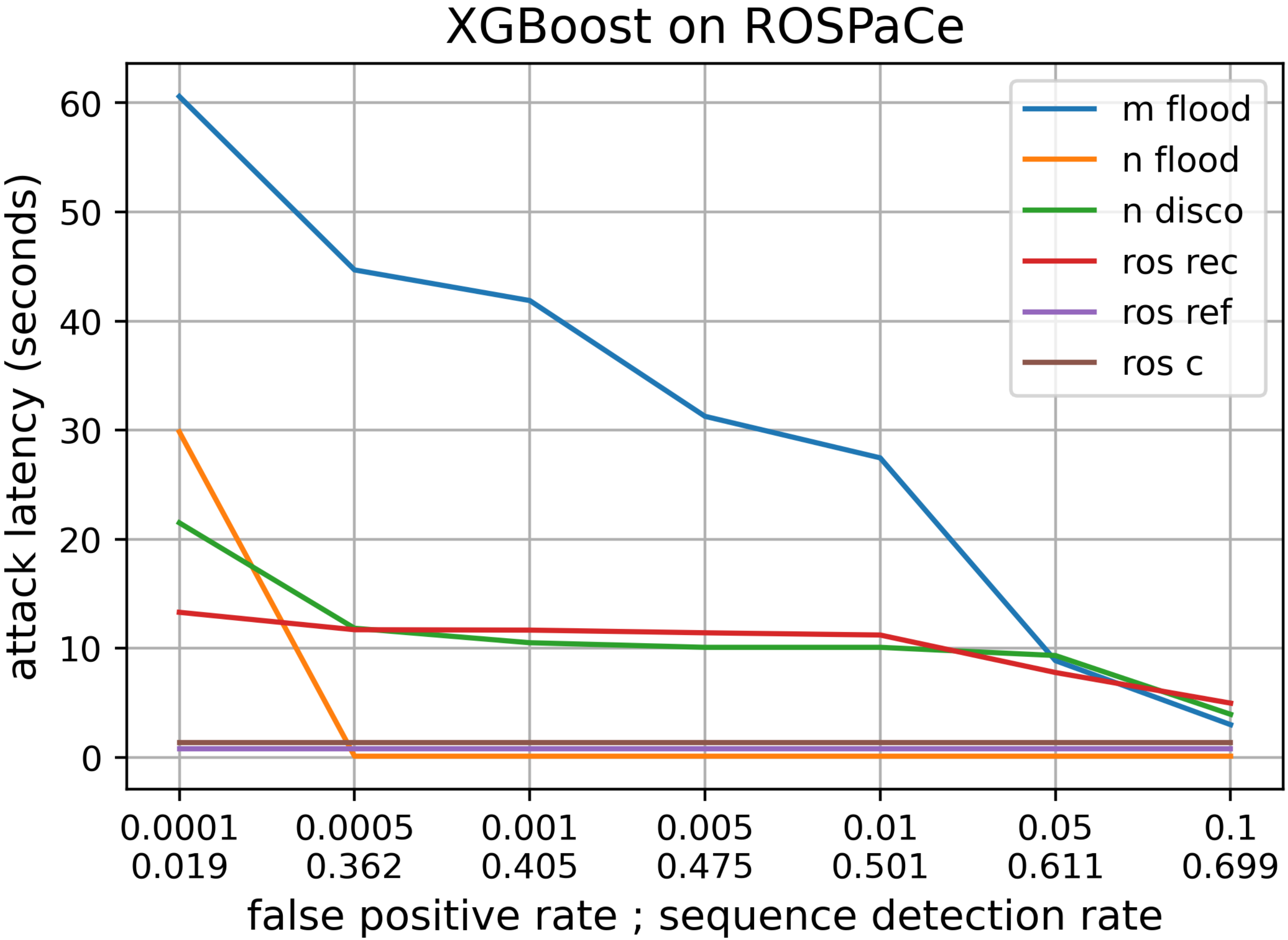
Some results: with «traditional» metrics

XGBOOST			LSTM CD		
Accuracy	Recall	F1	Accuracy	Recall	F1
0.927	0.991	0.952	0.879	0.911	0.953





What about average latency?



Presentation Outline

1. Context: why and how anomaly-based intrusion detection

2. Which classifier
   – The role of DNNs
   – Detection of unknown attacks (zero-days)
   – Take advantage of many: stacking

3. A forgotten measure: attack latency

4. What's next: defend against Advanced Persistent Threats

Advanced Persistent Threats

Advanced, well-financed attack campaign with a full spectrum of intelligence-gathering techniques.

Persistent, from highly determined and persistent attackers. One of the attackers' goals is maintaining long-term access to the target.

Threats executed by coordinated human actions rather than mindless automated code.

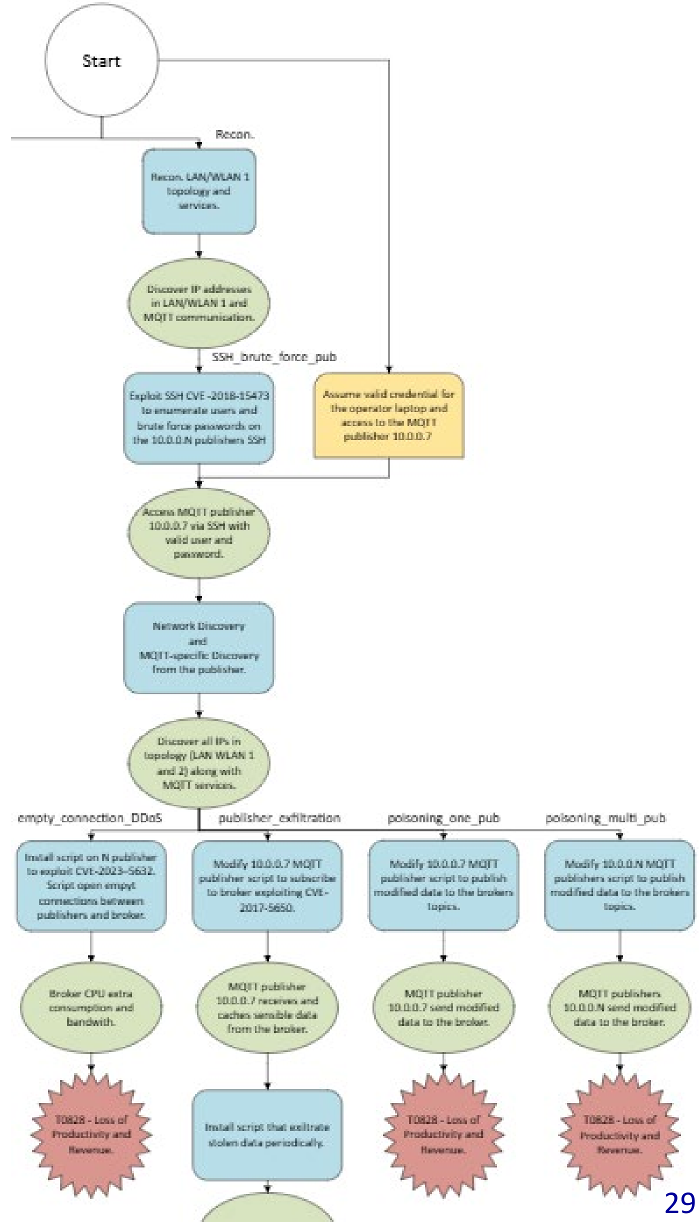
Anomaly detectors for APTs

A shift of perspective:

- not just «detect an attack»,

  but

- interrupt the attack path before the goal is reached



What is missing with respect to everything we have seen:

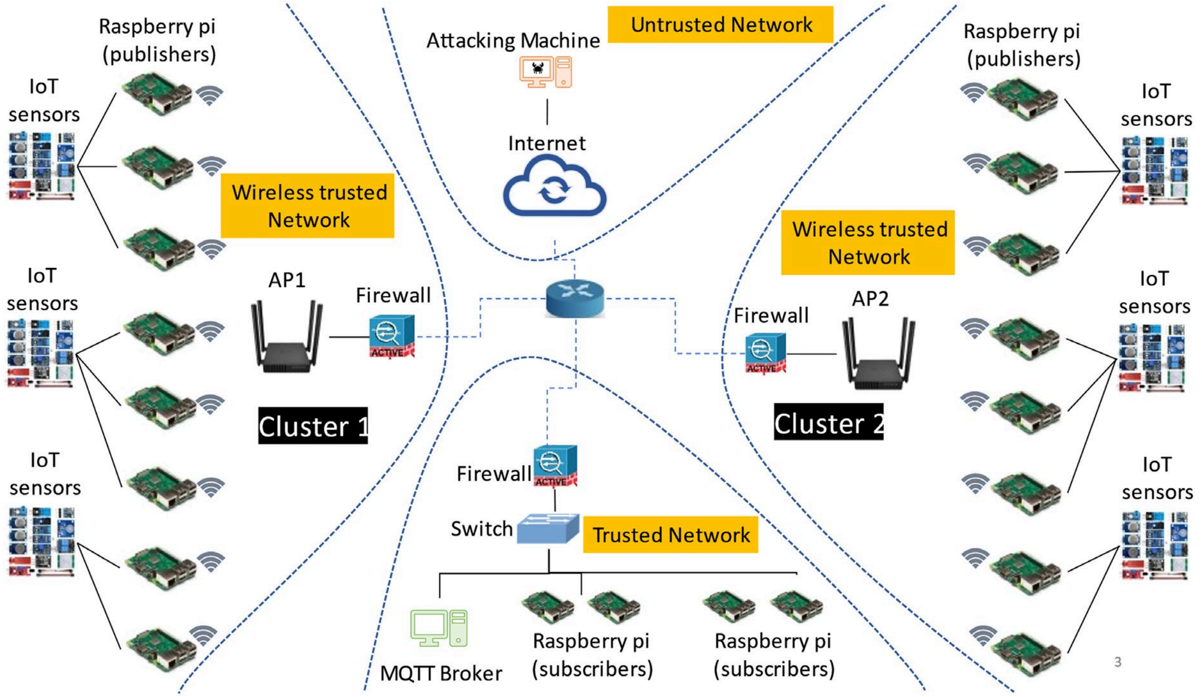
- Above all, datasets!
- Then, algorithms for time series exists (even if *maybe* not so much applied to IDS *yet*)

Let’s try to build a dataset

Industrial network traffic dataset DoS/DDoS-MQTT-IoT (publish/subscribe)

Simulate Network environment using DDoShield-IoT

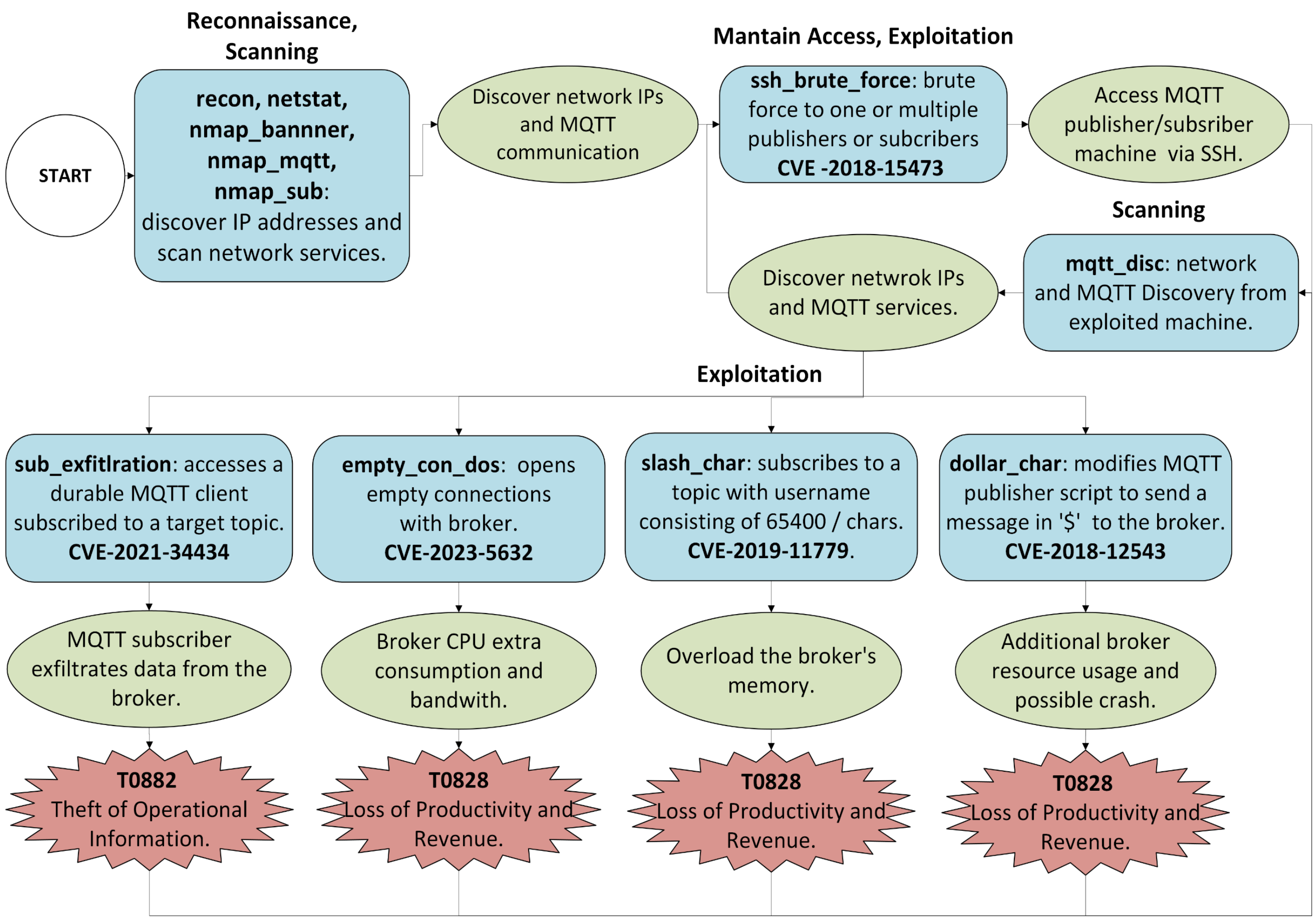
- Can replay dataset .pcap file and simulate network normal behavior **<- and we can craft attack!**



Alatram, Alaa, et al. "DoS/DDoS-MQTT-IoT: A dataset for evaluating intrusions in IoT networks using the MQTT protocol." *Computer Networks* 231 (2023): 109809.

De Vivo, Simona, et al. "DDoShield-IoT: A Testbed for Simulating and Lightweight Detection of IoT Botnet DDoS Attacks." *2024 54th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*. IEEE, 2024.

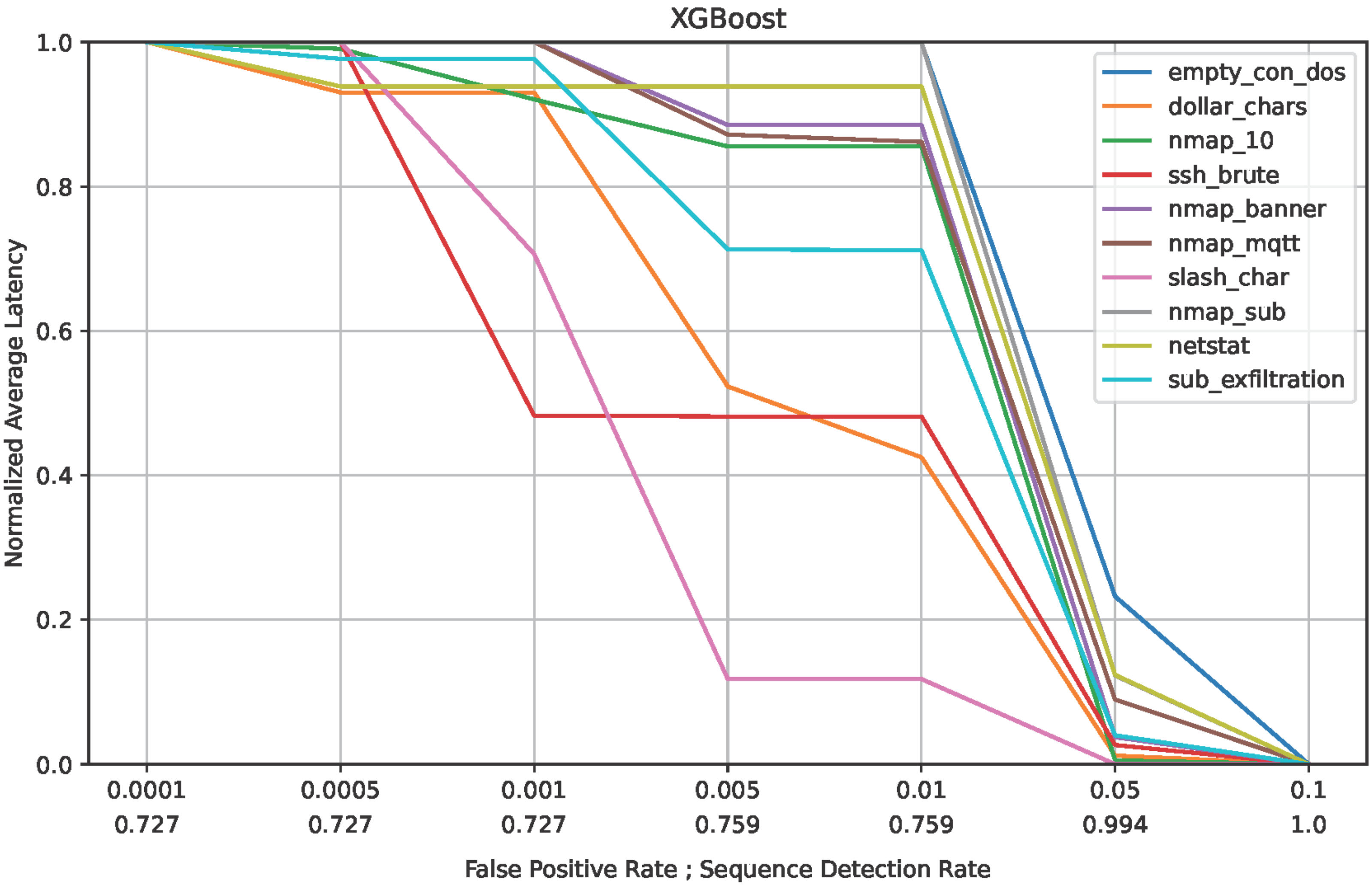
Design and implement the attack paths



MUR FLEGREA - *Federated Learning for Generative Emulation of Advanced Persistent Threats*

Train-test; analyze results

not good but just our first try



(Finally!) Wrapping Up…

(Finally!) Wrapping Up…

Anomaly-based IDS

- (only?) alternative to the signature/rule-based model
- Promising against unknowns

Not easy to deploy/customize

- Target-specific attack datasets needed!

And worst yet to come?

- APT as the new challenge to IDSs