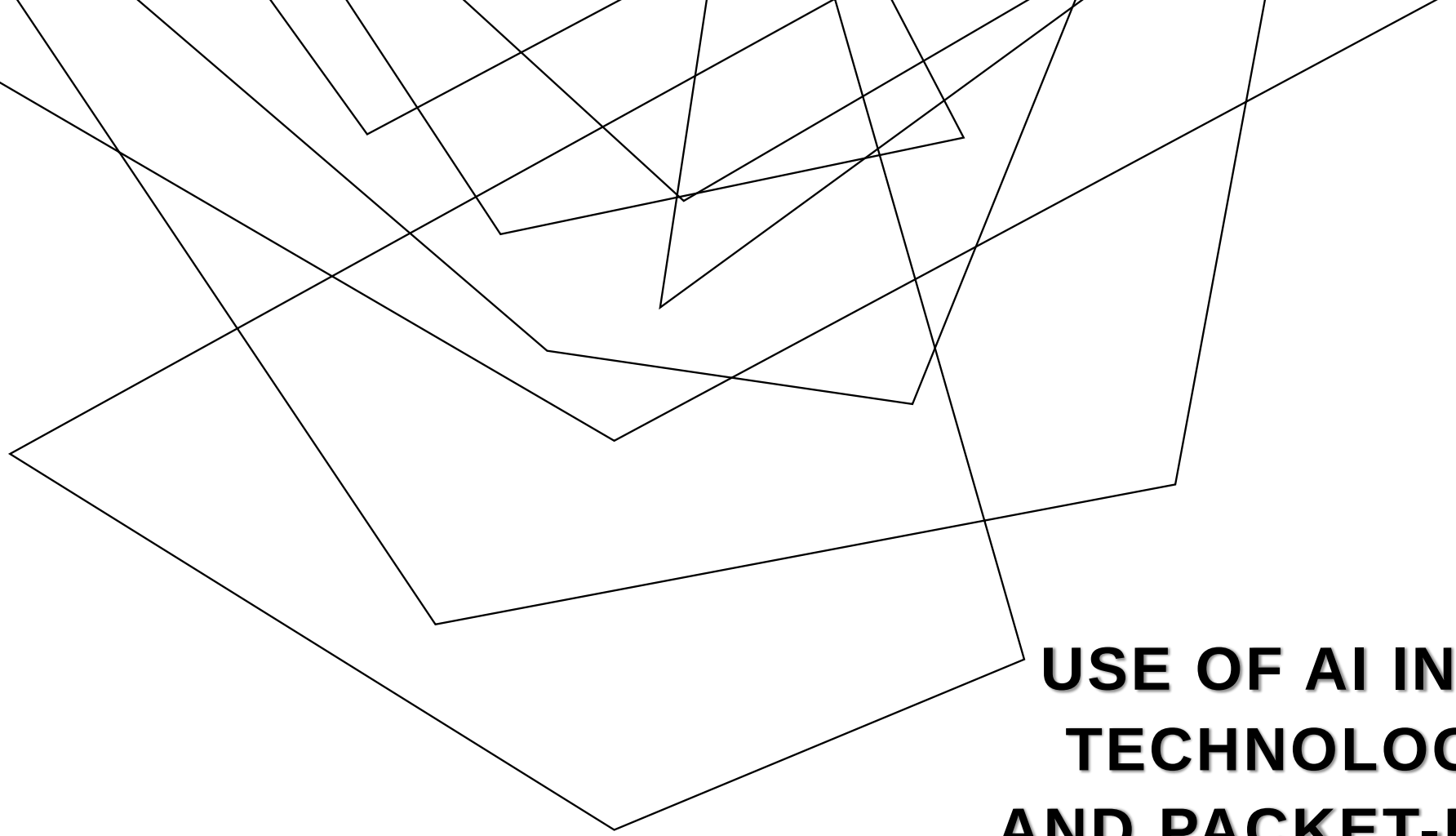


USE OF AI IN OPERATIONAL TECHNOLOGY NETWORKS AND PACKET-BASED ATTACKS DETECTION

Zoltán Dobrády, Szilárd László Takács, Timót Hidvegi

ABOUT US

The name of our research team:

„Industrial and Research Lab for Cybersecurity”

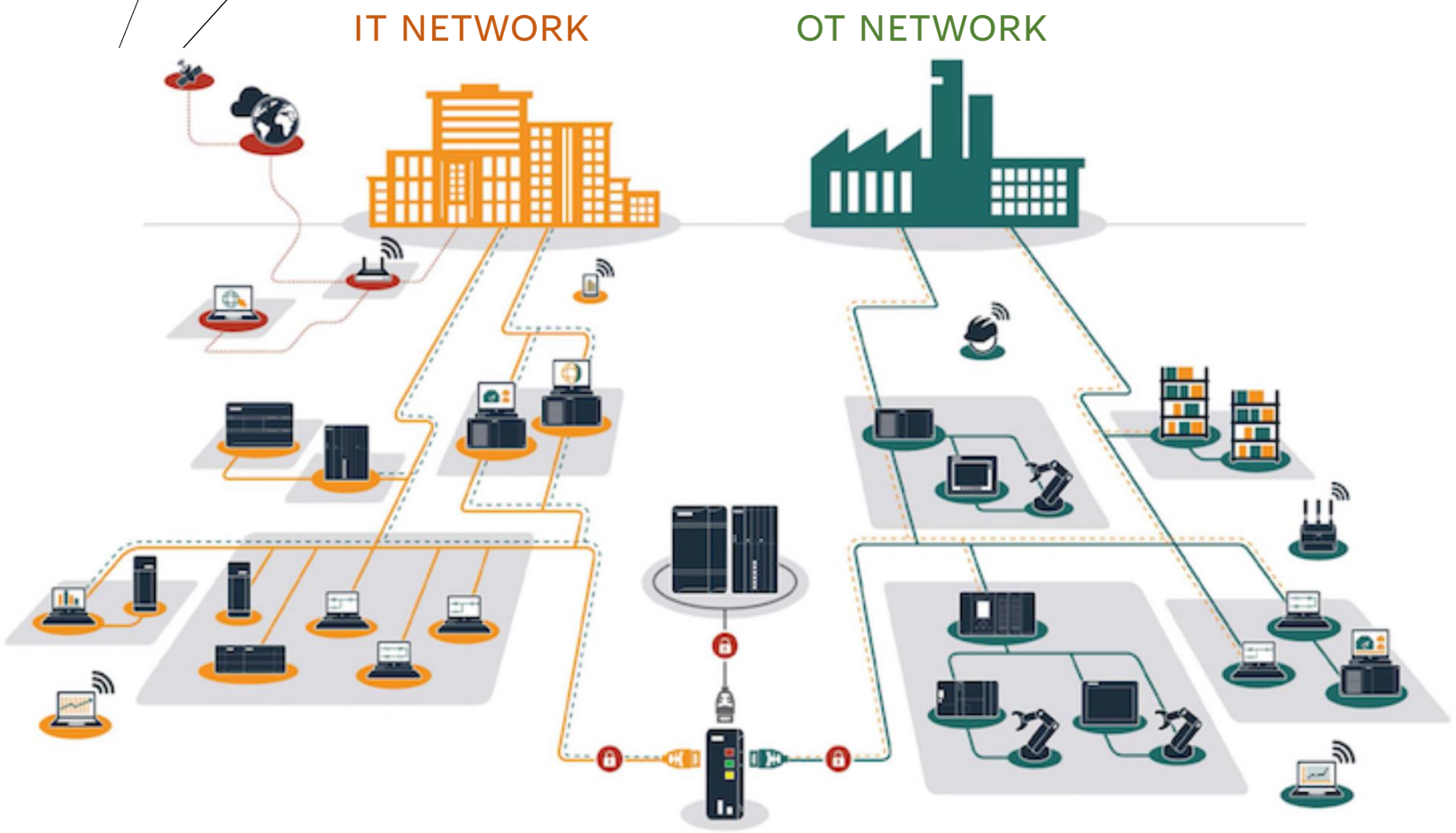
Domain:

https://cyberseclab.eu/

Research area:

Artificial intelligence applied to automotive and industrial systems.

NETWORK OVERVIEW



Source: inprosec

- An OT (Operational Technology) network refers to a specific type of network used in industrial and critical infrastructure environments.
- This is the communication between the machine (PLC) units
- Includes communication of peripherals (e.g. ProfiNET, ProfiBUS, ect.) to the machine
- Our research analyses data packets on this network

PROBLEM

VULNERABILITY

OT systems or networks may be susceptible to targeted cyberattacks
Contrary to IT systems, these peripherals are not intended to ward off cyberattacks

PRODUCTION DOWNTIME

Current example: Recently one big car manufacture acknowledged to suffer a targeted cyberattack witch caused the complete stop off production lines across Europe.

DEFECTIVE PRODUCT MANUFACTURING

Even one manipulated data packet is enough to cause a peripheral to malfunction. This leads to the production of defective products, which causes a lot of financial costs.

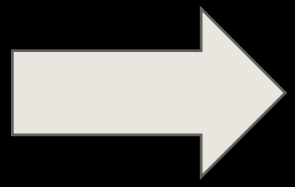
DATA PACKET ANALYSIS

Currently this data packet analysis is done manually after the attack with filtering methods(e.g. wireshark, tshark, ect.) which poses an additional error source, and it takes a great deal of time

POSSIBLE SOLUTION

„Real time” analysis

It is imperative that the packets are analyzed in real time, rather than after the attack has occurred.

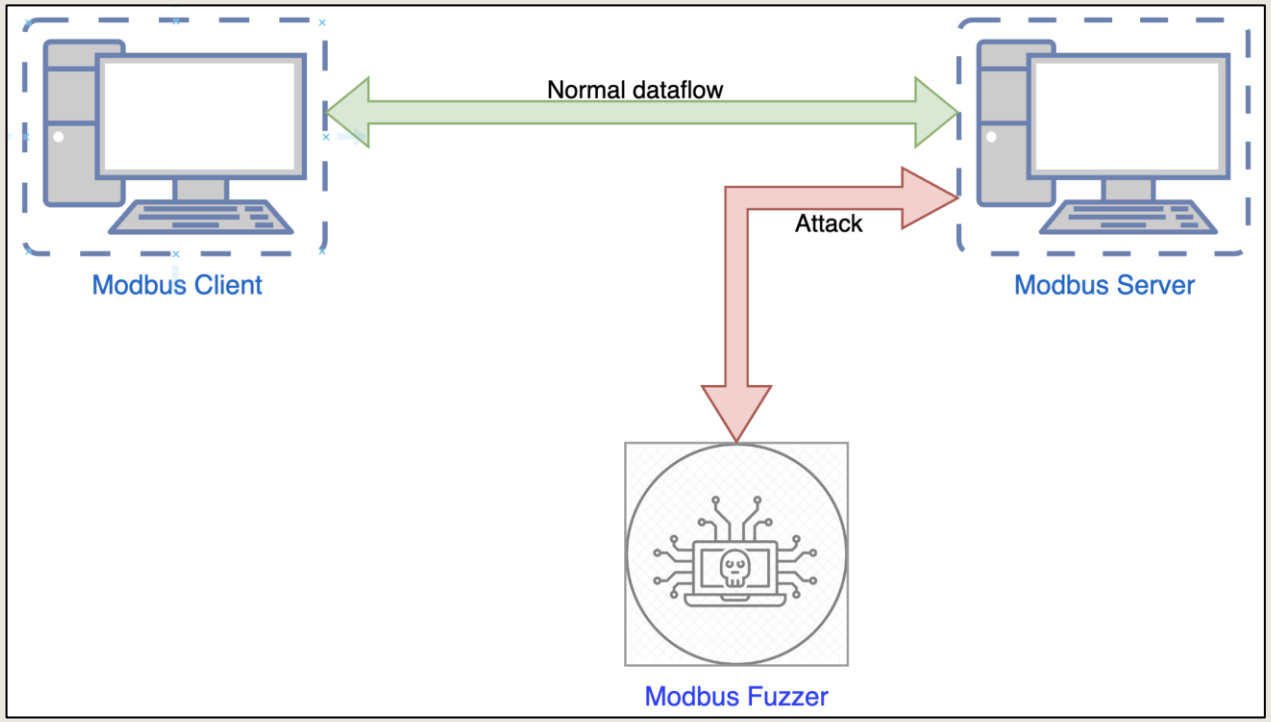


Use of Artificial intelligent

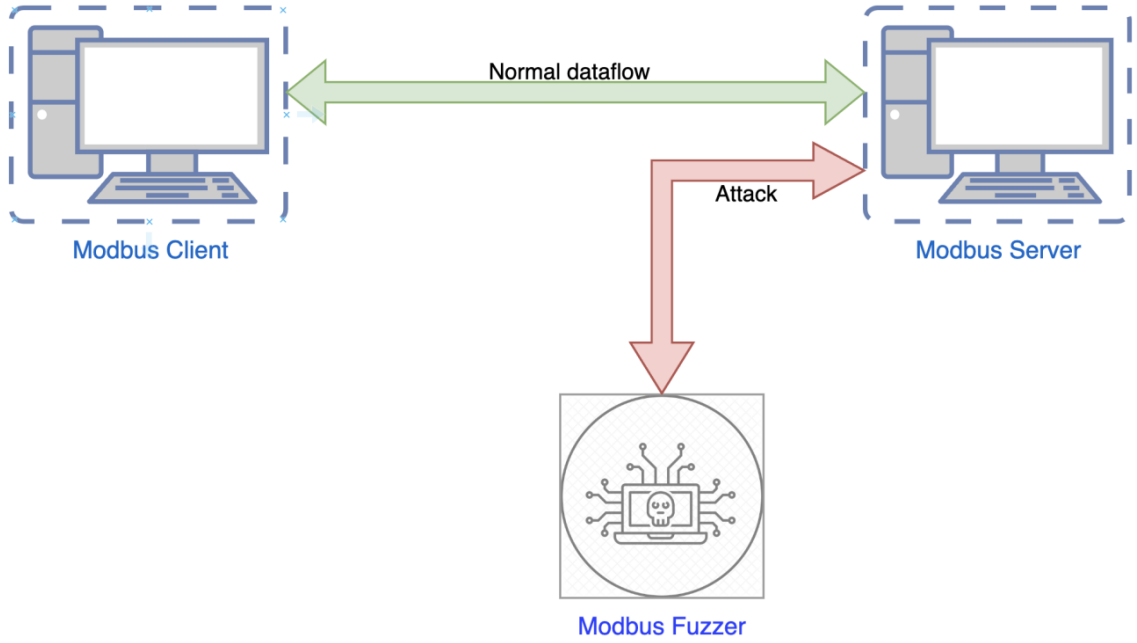
After selecting the appropriate learning method for the A.I., packet analysis can be used to predict attacks

ATTACKING MECHANISM

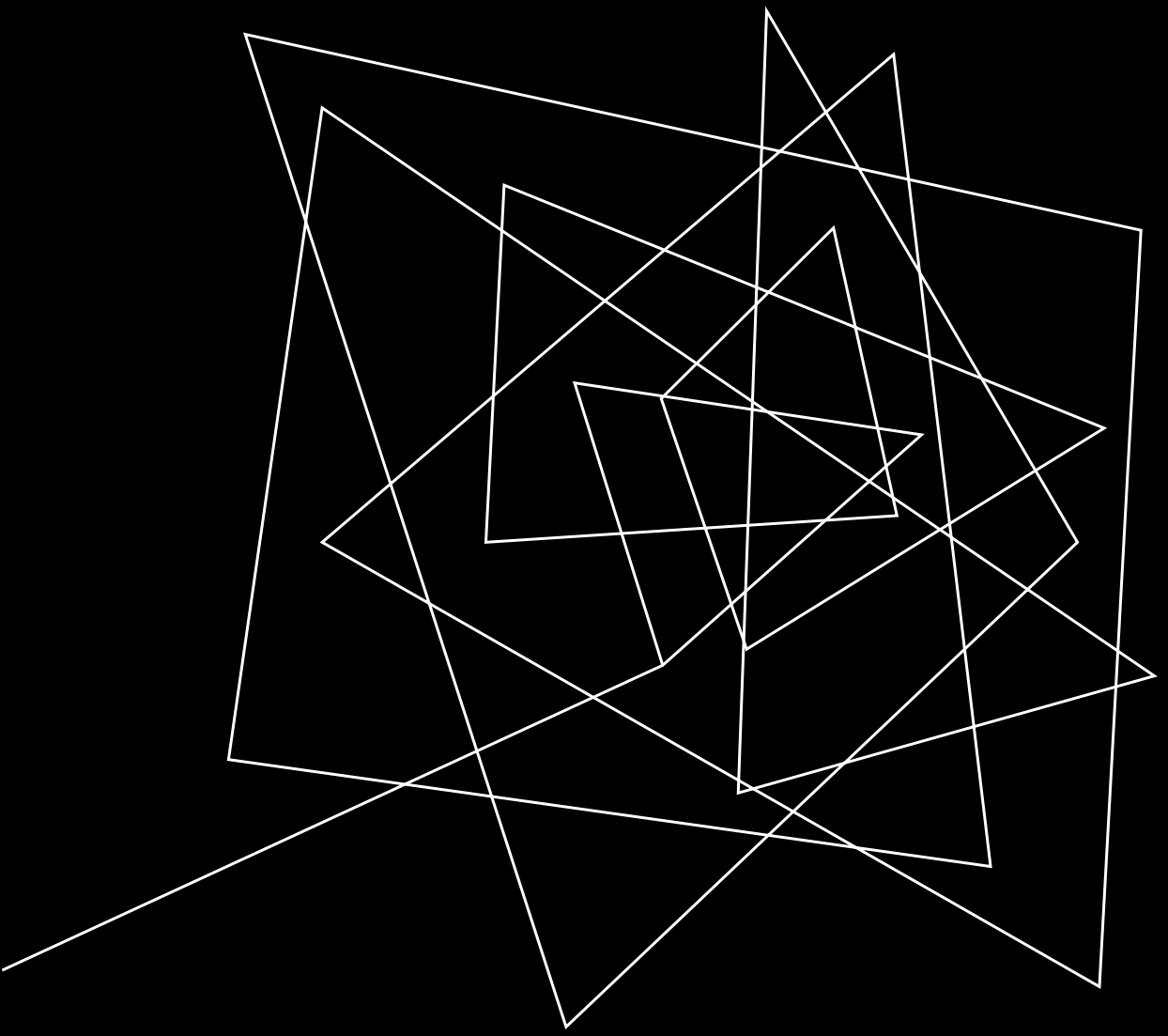
- Flooding the data stream with large amounts of data
- In response the server is left with little or no resource for communication to the client



VIRTUAL TESTING ENVIRONMENT



- Used three virtual machines
- Standard Modbus Server/Client communication
- Using a fuzzer, we executed a cyber attack by generating continuous requests across the network
- Fuzzer: generally used for examining and testing applications/protocols.



A.I.

- Artificial Intelligence, refers to the simulation of human intelligence in machines that are programmed to think and learn like humans.

- We used lot of A.I. methods:
  - Statistical learning
  - Natural Language processing

PREPARING GENERATED DATA FOR A.I.

- The generated data packets had to be labelled
- Irrelevant columns for the A.I. have been removed.
- For the natural learning model we used word-based tokenization on the dataset.

Before:

	Source	Destination	Protocol	Length	Info
0	PcsCompu_22:46:4f	Broadcast	ARP	60	Who has 192.168.56.113? Tell 192.168.56.114
1	PcsCompu_75:69:b0	PcsCompu_22:46:4f	ARP	42	192.168.56.113 is at 08:00:27:75:69:b0
2	192.168.56.114	192.168.56.113	TCP	74	36610 > 502 [SYN] Seq=0 Win=64240 Len=0 MSS=...
3	192.168.56.113	192.168.56.114	TCP	74	502 > 36610 [SYN, ACK] Seq=0 Ack=1 Win=65160...
4	192.168.56.114	192.168.56.113	TCP	66	36610 > 502 [ACK] Seq=1 Ack=1 Win=64256 Len=...
...	...	...	...	...	...
430913	13.229.250.8	192.168.56.113	TCP	54	1234 > 502 [SYN] Seq=0 Win=8192 Len=0
430914	224.168.77.124	192.168.56.113	TCP	54	1234 > 502 [SYN] Seq=0 Win=8192 Len=0
430915	130.136.216.64	192.168.56.113	TCP	54	1234 > 502 [SYN] Seq=0 Win=8192 Len=0
430916	82.246.78.254	192.168.56.113	TCP	54	1234 > 502 [SYN] Seq=0 Win=8192 Len=0
430917	139.102.116.152	192.168.56.113	TCP	54	1234 > 502 [SYN] Seq=0 Win=8192 Len=0

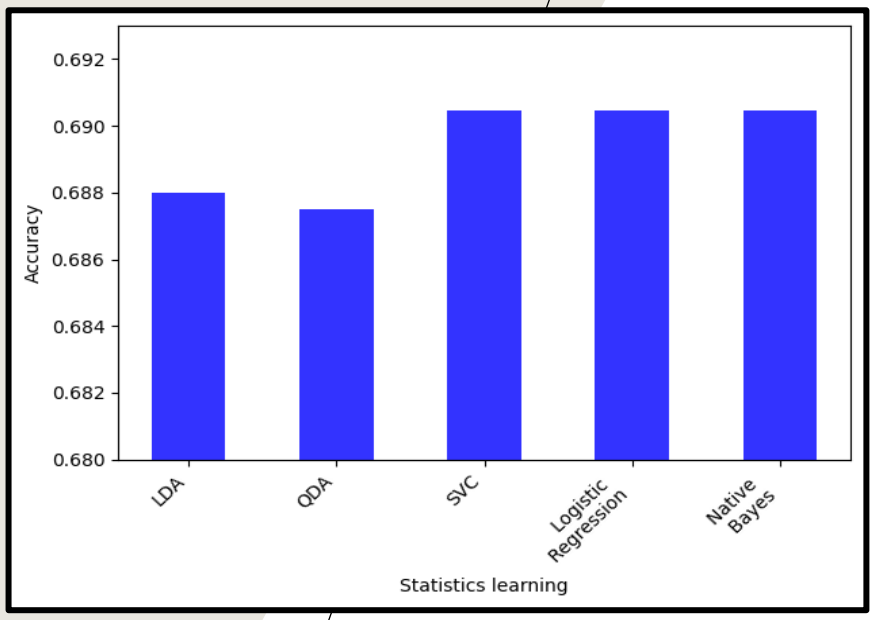
430918 rows × 6 columns

After:

Data				IsAttack?
Source IP	Destination IP	Protocol	Length	(0 = No attack,1 = Attack)
210.11.140.185	192.168.56.113	TCP	54	1
14.221.153.215	192.168.56.113	TCP	54	1
176.137.215.247	192.168.56.113	TCP	54	1
88.64.227.9 192	192.168.56.113	TCP	60	0
3.156.6.135 192	192.168.56.113	TCP	60	0

STATISTICAL LEARNING MODEL RESULTS

Models	Accuracy score
Linear regression	0.443869
Logistic regression	69.04627
kNN(k Nearest Neighbors)	31.23024
Linear Discriminant Analysis	68.80032
Quadratic Discriminant Analysis	68.75143
Support Vector Machine	69.04627
Naive Bayes	69.04627
Random Forest	46.13880



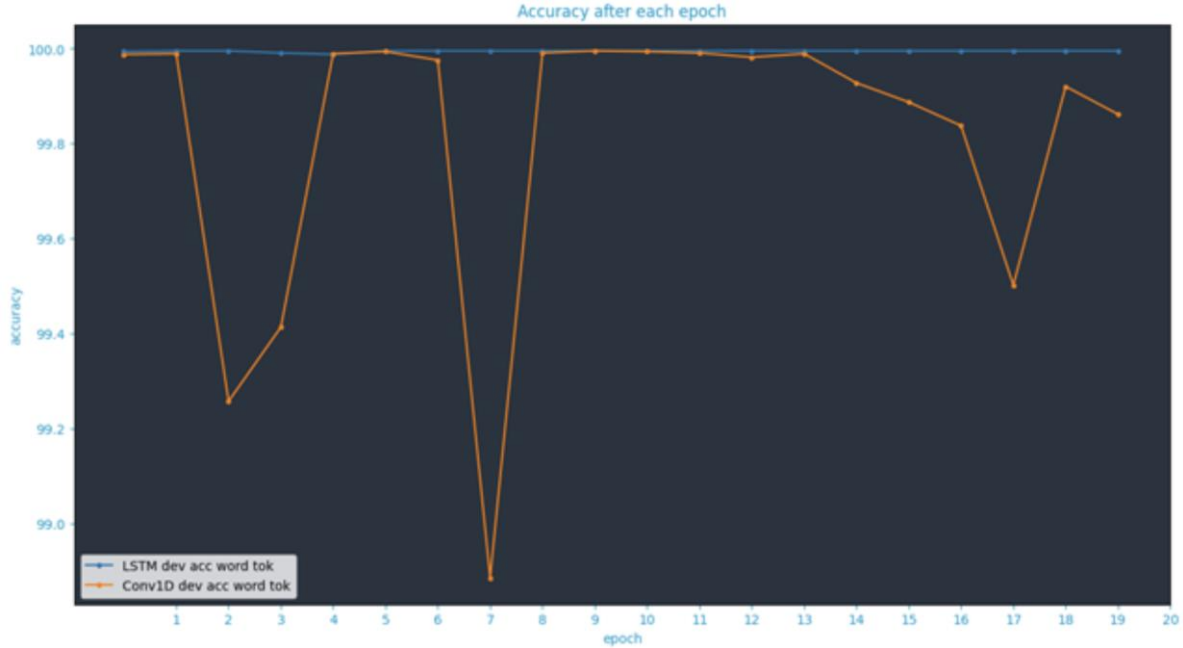
THE MODELS:

- REFER TO ALGORITHMS AND ARCHITECTURES DESIGNED TO PERFORM SPECIFIC TASKS OR SOLVE PARTICULAR PROBLEMS.
- THESE MODELS ARE CREATED THROUGH MACHINE LEARNING TECHNIQUES
- TRAINED ON LARGE DATASETS TO RECOGNIZE PATTERNS AND MAKE PREDICTIONS

ACCURACY SCORE:

- REFERS TO A METRIC USED TO MEASURE THE PERFORMANCE OF A CLASSIFICATION ALGORITHM

RESULT AND ANALYSIS ON NATUREL LEARNING MODE



- TWO MODELS ARE USED???: WE USED LSTM AND 1D CONVOLUTION METODES
- THE MODELS WERE TRAINED OVER 20 EPOCHS
- CONTAINS NEURAL NETWORKS???
- THE RESULTS ARE TOO EXTREME
- THIS COULD BE BECAUSE WE WERE WORKING WITH MULTIPLE DATA SETS AND THERE MAY BE OVERFITTED

CONCLUSION

- Cyber security is an important issue that we need to improve.
- Machine learning is the optimal solution to detect attacks.
- Firstly, we apply statistical learning, which gives a detection accuracy about of 70%. Several statistical methods were tried, but Support Vector Machines proved to be the best.
- Secondly, we encode packet information into sentences and use natural language models including Long-Short Memory and 1D convolutions.
- In both cases, we assumed that the data was overfitted, since the models could only focus on the source IP address and ignored the rest of the data.
- In summary, natural language models may not be suitable for identifying attacks due to their overfitting nature. Statistical learning algorithms alone may not be accurate enough to detect attacks, but in combination with other algorithms, they can improve protection and optimize attack identification.

THANK YOU

Zoltán Dobrády

zoltan.dobrady@hotmail.com

Szilárd László Takács

szhilank@gmail.com

Timót Hidvégi

Timot.hidvegi@cyberseclab.eu