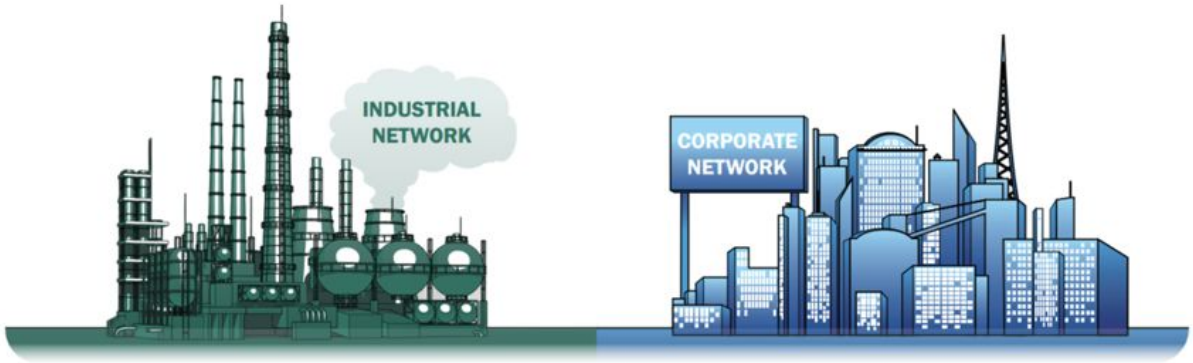
Network Anomaly Detection in Modbus TCP Industrial Control Systems

RP1 #52: Industrial Control Systems Research

Philipp Mieden & Rutger Beltman, 2020
Supervisor: Bartosz Czaszynski, Deloitte

Industrial Network VS Corporate Network

Different priorities



1. AVAILABILITY
2. INTEGRITY
3. CONFIDENTIALITY

1. CONFIDENTIALITY
2. INTEGRITY
3. AVAILABILITY

Problems for securing ICS networks

- Expensive hardware with long lifetime
- Many proprietary products with very little documentation available
- Licensing of a facility often prevents applying patches
- Availability: even small downtime impossible
- No security by default: no encryption, no authentication
- Devices not hardened: crash on ping etc

Countermeasures

- Network segmentation
- Intrusion Detection Systems / Monitoring
  - Strictly defined procedures, suitable for:
    - rule-based detection
    - **anomaly detection**

Research Questions

- How does malware look like on an ICS network?
- How does this differ from regular IT systems?
- Are pattern based / machine learning based solutions applicable?

Related Work

- Marthur et al. presents the Secure Water Treatment (SWaT) testbed for research on ICS security
- Goh et al. carried out a multitude of different attacks on SWaT with different attack types and created the SWaT Dataset
- Kravchick et al. tested two unsupervised machine learning methods on SWaT

Methodology

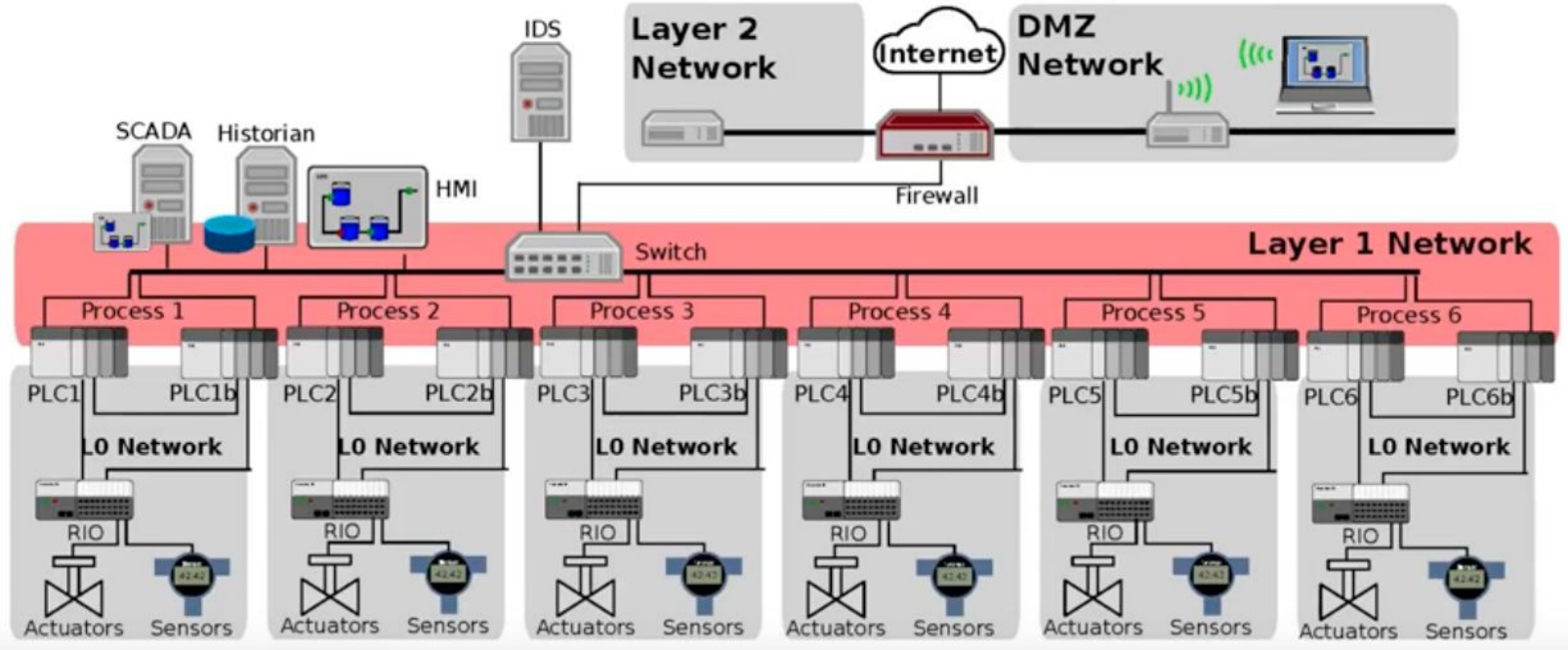
- Secure Water Treatment (SWaT) testbed dataset 2015 (100GB+ CSVs)
- Clean and encode the dataset to make it usable for the Deep Neural Network
- Train two different deep learning algorithms with Keras and Tensorflow
  - Sequential Dense DNN
  - Long Short Term Memory (LSTM) DNN

Dataset

- Secure Water Treatment (SWaT) from Singapore University of Technology and Design
  - Modern water treatment facility, with network segmentation
  - 6 Stage process: mechanical filtering and chemical cleaning
  - Well documented testbed
  - CSVs for Network and Physical data
  - Unmodified network captures in PCAP format
  - Evaluated in related research

Testbed

Testbed


IDS
Layer 2 Network
Internet
DMZ Network
SCADA
Historian
HMI
Firewall
Switch
Layer 1 Network
Process 1
Process 2
Process 3
Process 4
Process 5
Process 6
PLC1
PLC1b
PLC2
PLC2b
PLC3
PLC3b
PLC4
PLC4b
PLC5
PLC5b
PLC6
PLC6b
L0 Network
RIO
Actuators
Sensors

Dataset Anatomy

Dec 2015	Jun 2017	Jul 2019	Dec 2019
- Attack Information - Network CSV - Labeled Physical CSV **EVALUATED**	400GB of PCAPs with benign traffic	- CSVs with Physical Sensor readings - no attack information	- 40GB of packet captures (normal operation + attacks) - Vague attack information **ANALYZED BUT NOT EVALUATED**

Devices

- PLC: Programmable Logic Controller(s), for controlling valves and pumps
- HMI: Human Management Interface(s), for displaying sensor values
- Engineer Workstation, for configuring PLCs
- Historian Server, for process monitoring

Attack Scenarios

- Single Stage Single Point (eg: open motorized valve to cause tank overflow)
- Single Stage Multi Point (eg: open valve and manipulate values on HMI)
- Multi Stage Single Point
- Multi Stage Multi Point

(Potential) Attack Impact

- Process Disruption
  - Tank Overflow
  - Motor / Pump Damage
- Process Manipulation?
  - Water throughput reduction
  - Causing failure to remove chemicals and hide it
    - Possible physical damage for humans

Attack Distribution

	AttackName	AttackType
75% TRAINING	Open MV-101	Single Stage Single Point
	Turn on P-102	Single Stage Single Point
	Increase by 1 mm every second	Single Stage Single Point
	Open MV-504	Single Stage Single Point
	Set value of AIT-202 as 6	Single Stage Single Point
	Water level increased above HH	Single Stage Single Point
	Set value of DPIT as >40kpa	Single Stage Single Point
	Set value of FIT-401 as <0.7	Single Stage Single Point
	Set value of FIT-401 as 0	Single Stage Single Point
	Close MV-304	Single Stage Single Point
	Do not let MV-303 open	Single Stage Single Point
	Decrease water level by 1mm each second	Single Stage Single Point
	Do not let MV-303 open	Single Stage Single Point
	Set value of AIT-504 to 16 uS/cm	Single Stage Single Point
	Set value of AIT-504 to 255 uS/cm	Single Stage Single Point
	Keep MV-101 on countinuosly; Value of LIT-101 set as 700 mm	Single Stage Multi Point
	Stop UV-401; Value of AIT502 set as 150; Force P-501 to remain on	Multi Stage Multi Point
	Value of DPIT-301 set to >0.4 bar; Keep MV-302 open; Keep P-602 closed	Multi Stage Multi Point
	Turn of P-203 and P-205	Single Stage Multi Point
	Set value of LIT-401 as 1000; P402 is kept on	Single Stage Multi Point
	P-101 is turned on continuosly; Set value of LIT-301 as 801 mm	Multi Stage Single Point
	Keep P-302 on contineoulsy; Value of LIT401 set as 600 mm till 1:26:01	Multi Stage Single Point
25% EVALUATION	Close P-302	Single Stage Single Point
	Turn on P-201; Turn on P-203; Turn on P-205	Single Stage Multi Point
	Turn P-101 on continuously; Turn MV-101 on continuously; Set value of LIT-101 as 700 mm; P-102 started itself because LIT301 level became low	Multi Stage Multi Point
	Set LIT-401 to less than L	Single Stage Single Point
	Set LIT-301 to above HH	Single Stage Single Point
	Set LIT-101 to above H	Single Stage Single Point
	Turn P-101 off	Single Stage Single Point
	Turn P-101 off; Keep P-102 off	Single Stage Multi Point
	Set LIT-101 to less than LL	Single Stage Single Point
	Close P-501; Set value of FIT-502 to 1.29 at 11:18:36	Single Stage Multi Point
	Set value of AIT402 as 260; Set value of AIT502 to 260	Multi Stage Single Point
	Set value of FIT-401 as 0.5; Set value of AIT-502 as 140 mV	Multi Stage Single Point
	Set value of FIT-401 as 0	Single Stage Single Point
	decrease value by 0.5 mm per second	Single Stage Single Point

Features

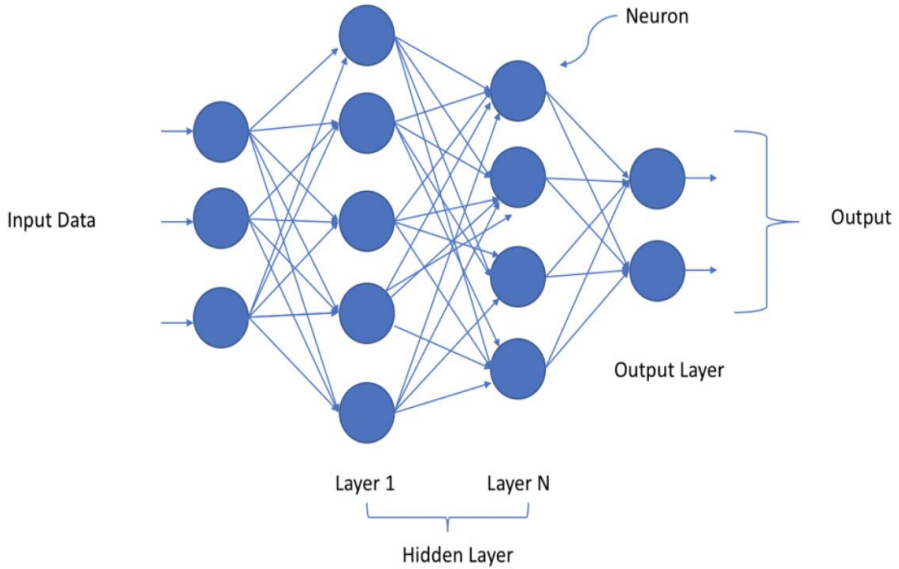
- 16 features in total
- IP address information
- Network Interface name and direction
- Protocol Name
- SCADA device tag
- Service Name and Port
- Modbus Function Code
- Modbus Transaction ID

Dataset Preprocessing

- Value encoding / normalization
  - strings: indexing
  - numeric values: z_score = (x - mean) / std
- Removal of columns that always contain unique values
  - Modbus_Value (modbus payload)
  - Sequence numbers
- UNIX Timestamp calculation based on Date and Time columns
- Labeling, mapping logic using attack timeframes and involved device addresses

Deep Neural Network (DNN)

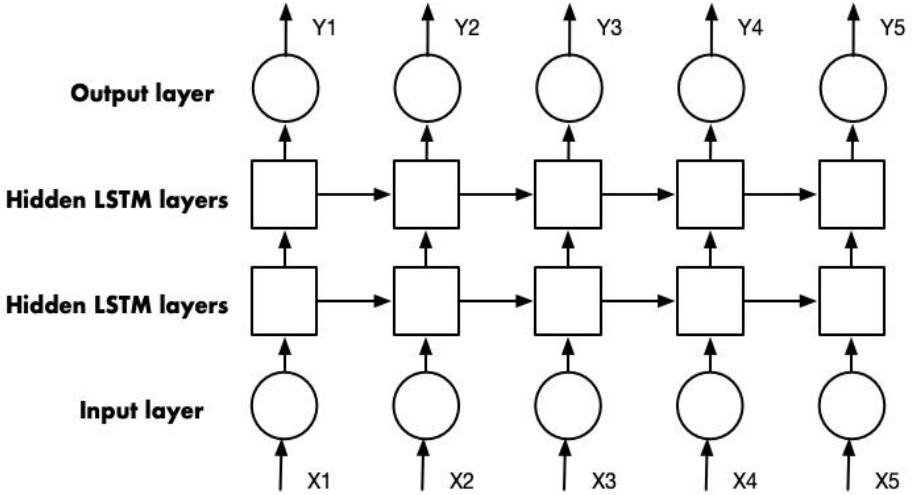
- Input layer with dimension of data
- N hidden layers
- Output Layer with the number of classes to predict (5 in our case: 1 normal, 4 attack types)



https://towardsdatascience.com/a-laymans-guide-to-deep-neural-networks-ddcea24847fb

Long Short Term Memory (LSTM) DNN

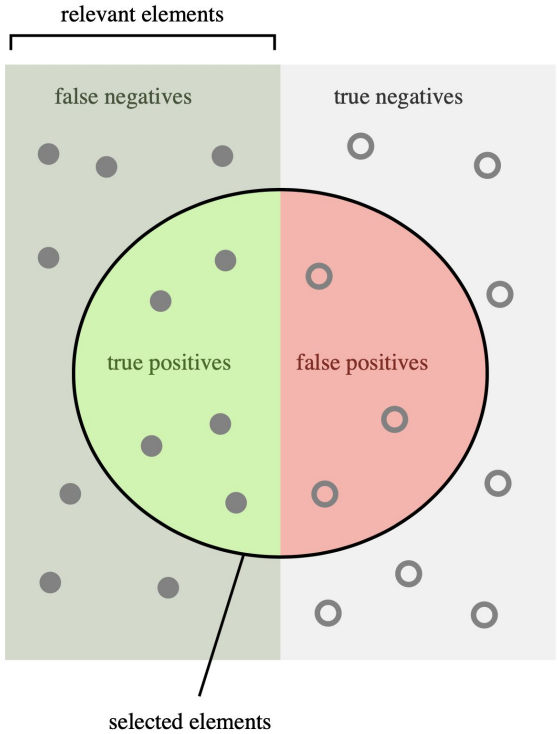
- Suited for time series data
- Increased training time
- Activation functions: softmax, relu
    - Problem: ReLU treats all negative values as 0, addressed via LeakyReLU



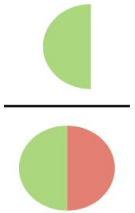
Challenges

- Dataset cleaning: Typos, typos, typos, missing data...
- Labeling: Network CSV not labeled
  - Attack information needed to be aggregated
- DNN configuration
- Hyperparameter tuning

Metrics



How many selected items are relevant?

Precision = 

How many relevant items are selected?

Recall = 

https://en.wikipedia.org/wiki/F1_score

Metrics

- F1 Score: Harmonic Mean between precision and recall
    - Useful to describe unbalanced data

$$F_1 = \left(\frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} \right) = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

Classification Results

- Experiments where the DNN would
  - exclusively predict one single class.
  - predict between normal and one other attack type

Experiment Results - DNN

Experiment #	Attack type	f1-score
1	SSSP	0.094
2	MSSP	0.005
3	SSSP	0.043
4	SSSP	0.083
5	SSSP	0.132
6	SSSP	0.200
7	SSSP	0.035

Experiment Results - LSTM

Experiment #	Attack type	f1-score
1	SSSP	0.063
2	SSSP	0.153
3	SSSP	0.133
4	SSSP	0.124
5	SSSP	0.016
6	SSSP	0.108
6	MSSP	0.025

Research Questions

- **How does malware look like on an ICS network?**
  - Infection and lateral movement are comparable to corporate networks
  - Common network protocols: Ethernet, IP, TCP, UDP, HTTP(S)
  - Targeting horribly outdated Windows workstations
    - Or PLCs that are (accidentally?) exposed to the internet

Research Questions

- **How does this differ from regular IT systems?**
  - For causing physical damage / process interruption: knowledge of domain specific protocols (CIP, ModBus, etc) and hardware
  - But more important: Knowledge about the physical process
    - Requires reconnaissance, to gather design documents etc
  - Objective:
    - Intellectual Property Theft
    - Cyber Warfare

Research Questions

- **Are pattern based / machine learning based solutions applicable?**
  - Yes, but need to be carefully adjusted
  - Still rely on human supervision
    - Potentially high alert frequency
    - Potentially high ratio of false positives

Conclusion

- LSTM DNN applicable
  - increased training time
- Multiclass classification for attack types difficult
  - requires sufficient amount of well suited training data
- Detecting an intruder in his early stages of lateral movement and reconnaissance can prevent further damage
- Detecting changes in the physical state of the plant?
  - If that happens, it’s already too late!

Conclusion

- Different priorities, but similar technologies
- Anatomy of an intrusion is identical
  - Common Network Intrusion Detection Systems can be deployed
    - But need parsing support for ICS protocols: Modbus, ENIP, CIP ...

Discussion

- How to make alert decisions understandable for a humans?
  - DNN == Blackbox
  - Ensemble Learning Methods for increased decision transparency?
    - Voting model
- DNN configuration
  - layer types / neurons
  - hyperparameters
  - optimizers
  - activation functions

Discussion

- Not every anomaly is an attack!
- Attacks may affect normal system behavior
  - more alerts / anomalies
- Even when detecting only parts of a malicious stream as anomalous
  - alert can reveal suspicious activity anyways
- High data volume from packet-based records
  - use summary structures? Events etc?

Future work

- Use MODBUS payload data for feature engineering
- Compare to unsupervised methods
- Attempt to encode certain columns with multi-hot encoding
- Hyper parameter optimization
- Feature extraction, eg: Principal Component Analysis (PCA)
- Run each experiments multiple time to get an average and standard deviation of all statistics

Experiment Results - DNN

Experiment #	Attack type	precision	recall	f1-score
1	SSSP	0.053	0.415	0.094
2	MSSP	0.003	0.033	0.005
3	SSSP	0.029	0.081	0.043
4	SSSP	0.047	0.355	0.083
5	SSSP	0.079	0.404	0.132
6	SSSP	0.143	0.334	0.200
7	SSSP	0.050	0.027	0.035

Experiment Results - LSTM

Experiment #	Attack type	precision	recall	f1-score
1	SSSP	0.036	0.267	0.063
2	SSSP	0.087	0.646	0.153
3	SSSP	0.130	0.136	0.133
4	SSSP	0.092	0.191	0.124
5	SSSP	0.111	0.009	0.016
6	SSSP	0.060	0.583	0.108
6	MSSP	0.013	0.441	0.025