FACULTY
OF INFORMATION
TECHNOLOGY

Anomaly Detection of ICS Traffic Using Statistical Features
PDS project, academic year 2021/2022

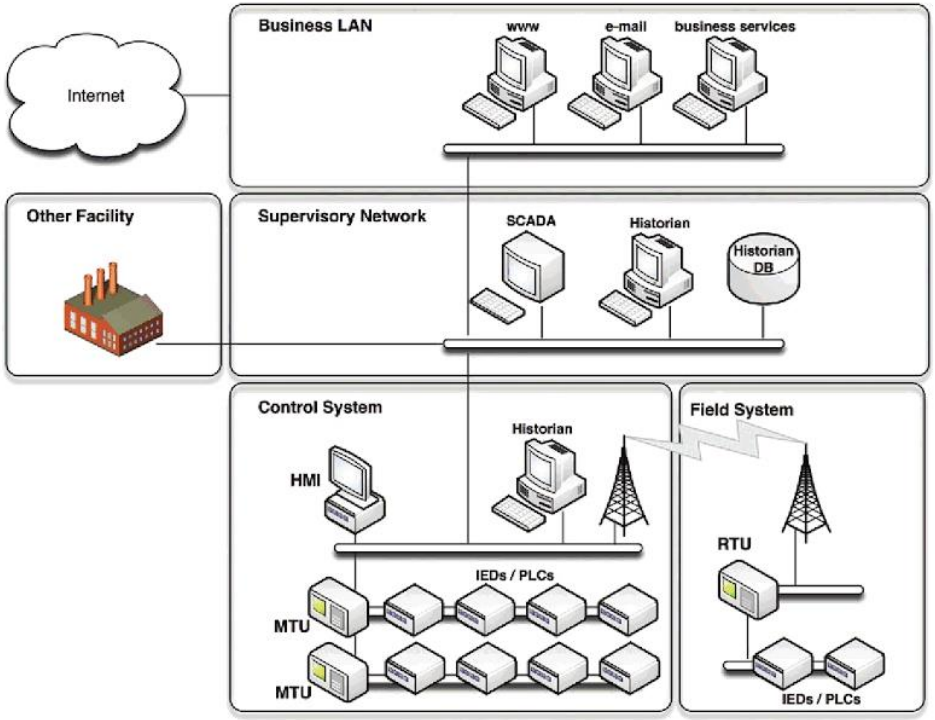
Dr. Petr Matoušek
Brno University of Technology, Faculty of Information Technology
Bozetechova 1/2, 612 66 Brno - Kralovo Pole
matousp@fit.vutbr.cz

PDS Project (for Czech students only)

- Goal: Implement anomaly detection of industrial traffic using statistical features.
- The project can be conducted through the following steps:
  1. Analyze IEC 104 communication, observe its communication patterns.
  2. Select statistical features that represent communication behavior.
  3. Process PCAP file(s) and extract relevant features from the traffic.
  4. Create a model for anomaly detection using a machine learning algorithm.
  5. Provide experiments with normal and anomalous data.
  6. Evaluate your results using FP, FN, accuracy, precision.
  7. Write the project report (see the recommended structure below).
  8. Submit the project (source codes + document) via FIT information system.
- Project deadline: 22nd April 2022
- Maximum points: 25 (extra points for extensions, see the last slide)
- Online consultations available using Forum in IS FIT.
- Individual project – each student creates its own solution.
- Plagiarism prohibited – see copyrights and the publication policy.

Industrial Control System (ICS) Communication

- Control and data transmission in industrial networks [1,2].



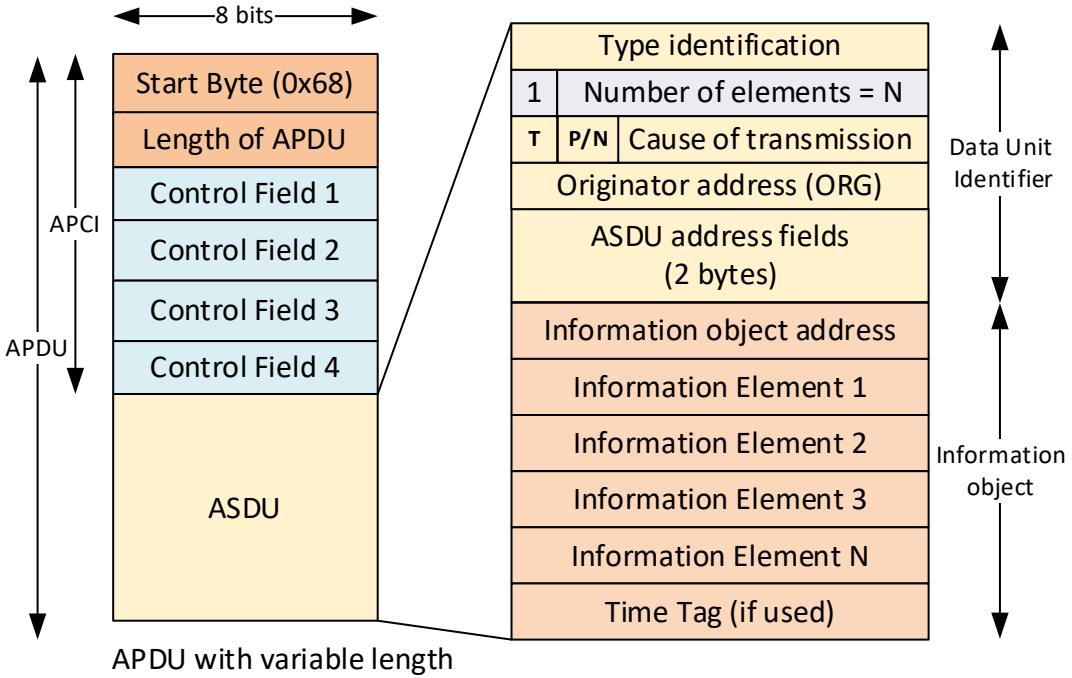
ICS Communication:

- Operational Technologies (OT) vs. Informational Technologies (IT).
- ICS systems build critical infrastructure (electricity, water, gas supply, traffic control, manufacturing processes, etc).
- Controls physical processes.
- Transmits data between end-points (IED, RTU, PLC) and control stations (HMI, servers).
- Employs industrial protocols.

[1] Trend Micro: Industrial Control System – Definition, See at https://www.trendmicro.com/vinfo/us/security/definition/industrial-control-system.
[2] K. Stouffer, S. Lightman, V. Pillitteri, M. Abrams, A. Hahn: Guide to Industrial Control Systems (ICS) Security, NIST SP 800-82 Rev. 2, May 2015.

Communication IEC 60870-5-104 (a.k.a. IEC 104)

- Control and monitoring communication in power grids and substations.
- Build upon TCP/IP.
- Each packet comprises an APDU or an APDU/ASDU, see the format [3].

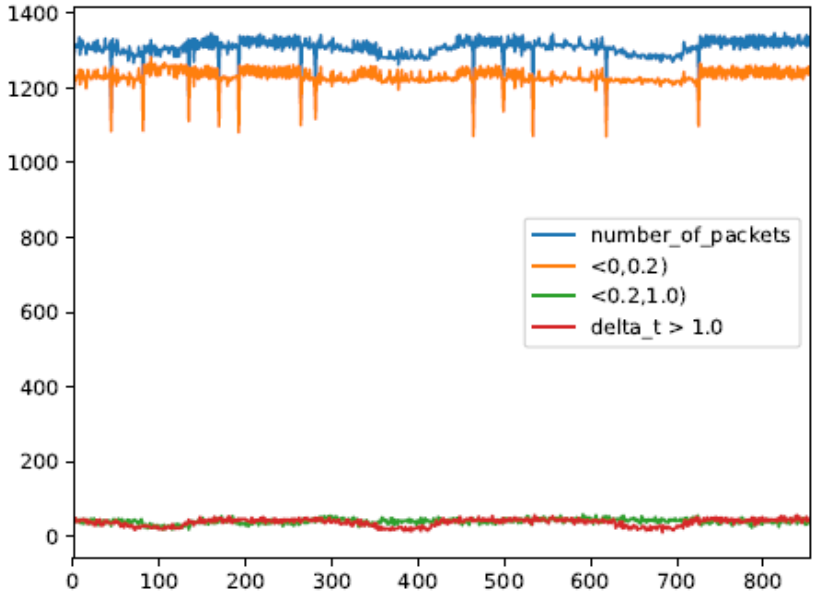


Type = 36 (M_ME_TF_1)		
1	Number of elements = 1	
T	P/N	COT=20 (interrogation)
Originator address = 1		
Common ASDU Address = 1		
IO Address (IOA) = 50		
Scaled Value (SVA) = 46		
Quality Descriptor (QDS) = 0x00		
CP56Time = Jul 11, 2018 16:23		

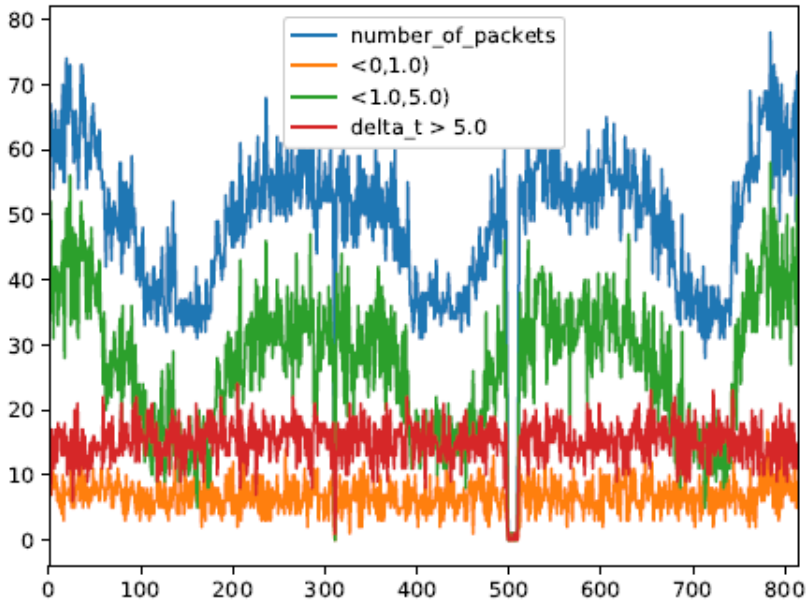
[3] MATOUŠEK Petr. Description and analysis of IEC 104 Protocol. FIT-TR-2017-12, Brno: FIT BUT, 2017.

Traffic patterns in industrial communication

- Machine to machine data exchange.
- Periodic behavior.
- Constant throughput.
- Master – slave(s) communication.
- Communication stability.



(a) Characteristics of the traffic from the master device.



[4] BURGETOVÁ Ivana and MATOUŠEK Petr. Statistical Methods for Anomaly Detection in Industrial Communication. IT-TR-2021-01, Brno, 2021.

1) Dataset analysis and feature selection

1. Open IEC 104 datasets provided in course repository in folder project/dataset. Analyze IEC 104 communication. Observe typical communication patterns [5,6].
2. Select interesting features (attributes) that sufficiently describe statistical behavior of IEC 104 communication, e.g., the number of transmitted bytes/packets, packet inter-arrival times, packet size, etc.
3. Extract selected features from IEC 104 packets and save them into a format suitable for machine learning, e.g., CSV, JSON, etc.
4. Analyze distribution of feature values within a dataset.

[5] R. R. R. Barbosa, R. Sadre and A. Pras, "A first look into SCADA network traffic," 2012 IEEE Network Operations and Management Symposium, 2012, pp. 518-521, doi: 10.1109/NOMS.2012.6211945.

[6] Valdes, Alfonso, and Steven Cheung. "Communication pattern anomaly detection in process control systems." In 2009 IEEE Conference on Technologies for Homeland Security, pp. 22-29. IEEE, 2009.

2) Building a model using statistical features

1. Choose a suitable machine learning technique for modeling statistical distribution of selected features, e.g.,
   - Simple statistical models using Box Plot (IQR) or Three-sigma rule [4],
   - One-class SVM classification (OC-SVM) [7],
   - Time Series and ARIMA model [8],
   - One-class Neural Networks (OC-NN) [9],
2. Find a tool/library to implement the model, e.g., scikit-learn.
3. Pre-process input data so that they fit the model.
4. Set initial model parameters based on the training dataset.
5. Define the threshold that separates normal and abnormal (anomalous) data.

[7] Lamrini, Bouchra & Gjini, Augustin & Daudin, Simon & Armando, François & Pratmarty, Pascal & Travé-Massuyès, Louise. (2018). Anomaly Detection Using Similarity-based One-Class SVM for Network Traffic Characterization.

[8] A. Lazaris and V. K. Prasanna, "Deep Learning Models For Aggregated Network Traffic Prediction," 2019 15th International Conference on Network and Service Management (CNSM), 2019, pp. 1-5, doi: 10.23919/CNSM46954.2019.9012669.

[9] Chalapathy, Raghavendra, Aditya Krishna Menon, and Sanjay Chawla. "Anomaly detection using one-class neural networks." arXiv preprint arXiv:1802.06360 (2018).

3) Making experiments to evaluate the model

1. Train the model with 2/3 of the training data. Use the last 1/3 of the data for validation (testing).
2. Observe accuracy of the model by computing the number of false positives and false negatives. Create confusion matrix to see the results.
3. If possible, improve your model to reach out higher precision.
4. (optional) Prepare a input set with anomalous data (missing packets, changed values, additional packets).
5. (optional) Test anomalous data with your model and evaluate anomaly detection.

4) Writing the report (5-10 pages)

Recommend document structure:

1. Problem description.
2. Description of IEC 104 dataset and interesting features.
3. Description of the anomaly detection method.
4. Implementation of data processing and building a model.
5. Testing and evaluation: experiments with extracted data, evaluation.
6. Discussion of the results.
7. Conclusion and contribution.

- Use BSc/MSc document template, see https://www.fit.vut.cz/study/theses/bachelor-theses/.

5) Project submission

1. Submit a zip file that includes following files (file *xlogin.zip*):
   - Readme.txt – your name, login, a list of files in the ZIP archive.
   - The project report in PDF format (file *xlogin.pdf*).
   - Source code of your scripts/tools you developer (optional).
   - Input data that you used to feed the model.
   - Output of your experiments.

Recommended references

- BURGETOVÁ Ivana, MATOUŠEK Petr and RYŠAVÝ Ondřej. Anomaly Detection of ICS Communication Using Statistical Models. In: *Proceedings of the 17th International Conference on Network Service Management (CNSM 2021)*. Izmir, 2021, pp. 166-172. 10.23919/CNSM52442.2021.9615510.

- Chih-Yuan Lin and Simin Nadjm-Tehrani. 2018. Understanding IEC-60870-5-104 Traffic Patterns in SCADA Networks. In Proceedings of the 4th ACM Workshop on Cyber-Physical System Security (CPSS '18). Association for Computing Machinery, New York, NY, USA, 51–60. DOI: 10.1145/3198458.3198460.

- S. V. B. Rakas, M. D. Stojanović and J. D. Marković-Petrović, "A Review of Research Work on Network-Based SCADA Intrusion Detection Systems," in *IEEE Access*, vol. 8, pp. 93083-93108, 2020, doi: 10.1109/ACCESS.2020.2994961.

- Christopher M. Bishop: *Pattern Recognition and Machine Learning*, Springer, 2006.

- Jiawei Han, Micheline Kamber, and Jian Pei. 2011. *Data Mining: Concepts and Techniques* (3rd. ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Concluding remarks

- The goal of the project is to analyze testing dataset and implement anomaly detection method based on statistical features using an existing method.
- A partial solution is also accepted. This must be explicitly stated in Readme.txt
- Any external tools, code, sources of information must be properly referenced, otherwise the work is considered as plagiarism.
- Extra points can be obtained for the following extensions:
  - Testing datasets with anomalies including description of the anomalies.
  - Application of advanced classification methods.
  - Modelling of high-level attributes, e.g., APDU format, ASDU type, CoT, etc.
  - Extra points can be given only when basic requirements are met.

Questions?