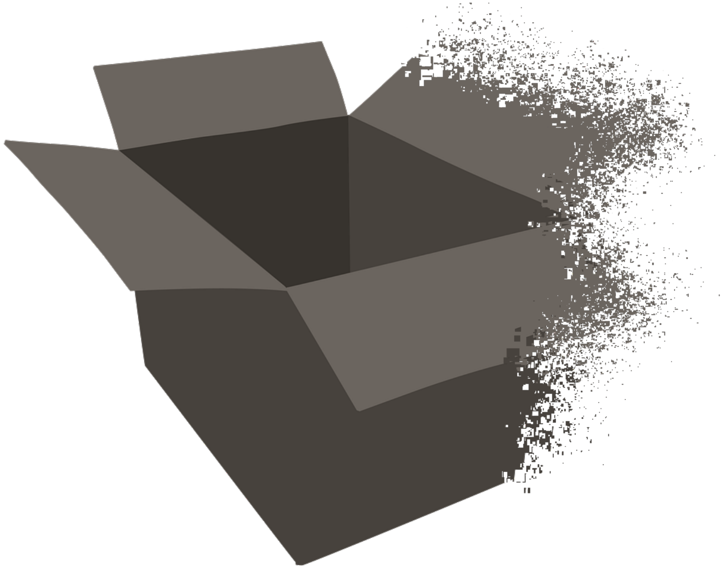


6th Nov 2020

Faculty 1: Chair of IT-Security | Master Thesis: Initial Talk

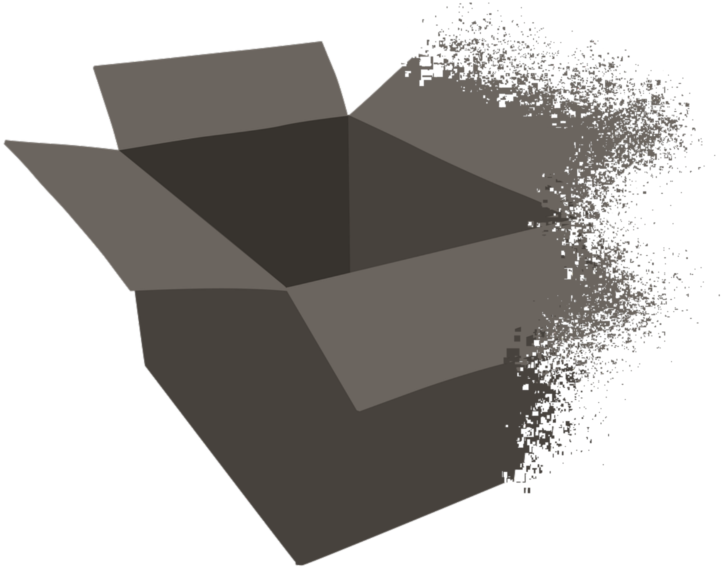
Unsupervised feature extraction from network traffic for content-based anomaly detection in industrial networks

B. Sc. Fabian Kopp

Supervisors:

M. Sc. Franka Schuster

Prof. Dr.-Ing. Andriy Panchenko



1 | Background: **Motivation**

- Cyber attacks like: *industroyer*, *blackengery*, *havex* or *stuxnet* are real threats
- Industrial field devices are *controlled and monitored* over computer networks
- Network based anomaly detection is one option to detect *0-day attacks*
- Feature extraction from network data is the *basis* for all detection algorithms
- *Unsupervised* machine learning methods can adapt to network protocols automatically

Brandenburg
University of Technology
Cottbus - Senftenberg

Network Data

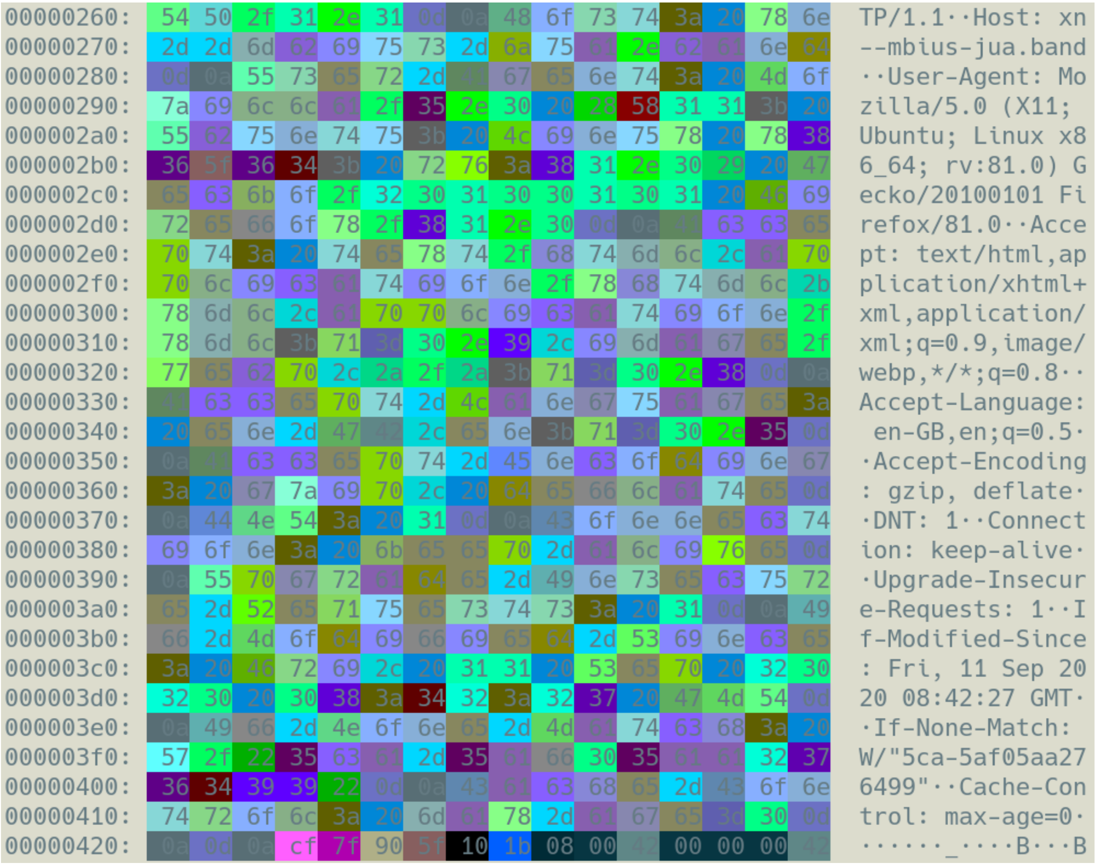
- Series of *packets*
- Packets consist of *layers*
- Layers have *header* and a *payload* field(s)
- Weak order of packets

Industrial Control Systems (*ICS*)

- Monitoring & Control functions
- Fixed network topology
- Proprietary Infrastructure

Intrusion Detection Systems (*IDS*)

- Used to secure networks
- Host deployment
    - Stack traces, log files, etc
- Network deployment
    - Raw data, aggregated traffic



Anomaly detection (*AD*)

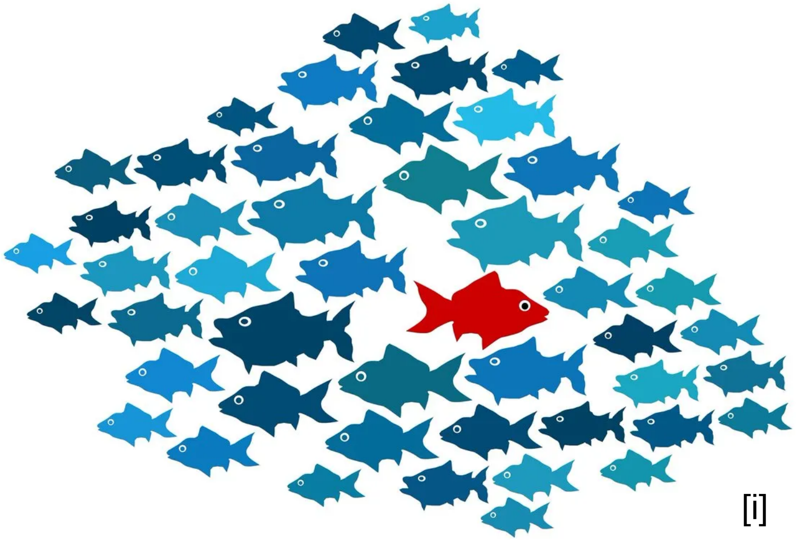
- Unsupervised *binary classification*
- *Learn* normal behavior - alert deviations
- Used in: fraud detection, intrusion detection

Types of anomalies ***[1]***

- Point → single payload (XSS)
- Collective → multiple payloads (Scan)
- Contextual → order of payloads (0-day)

Content based AD

- Detects intrusion on the byte level
- Payload and header information as basis



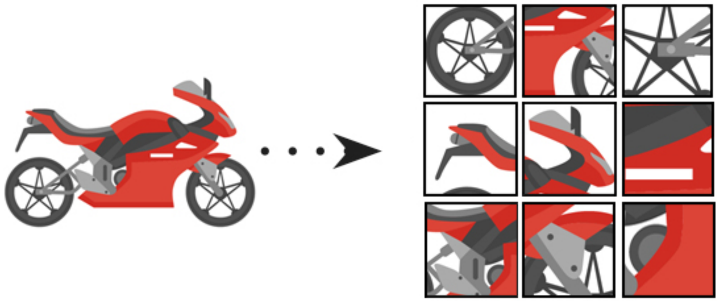
[i]

Feature extraction
- *Transformation* from input vectors to feature vectors

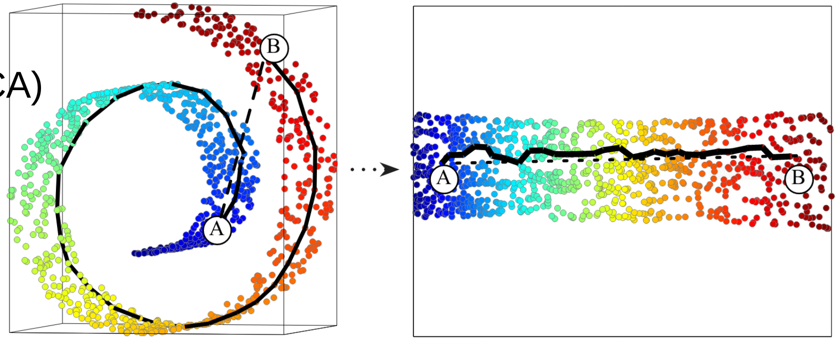
Feature selection
- Determine best subset of features

Feature engineering
- Consultation of *domain experts*

Representation learning
- Search for *model* which embeds feature vectors into subset
  - Linear methods
    - Principal component analysis (PCA)
  - Non-Linear methods
    - Kernel-PCA
    - Autoencoder



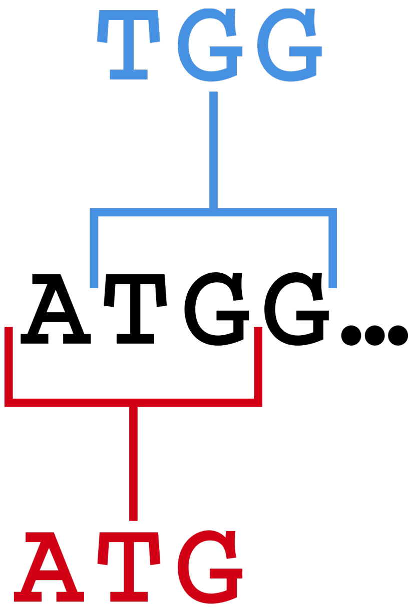
[ii]



[iii]

[2] ZOE: Content-based Anomaly Detection for Industrial Control Systems *(IEEE/IFIP DSN, 2018)*

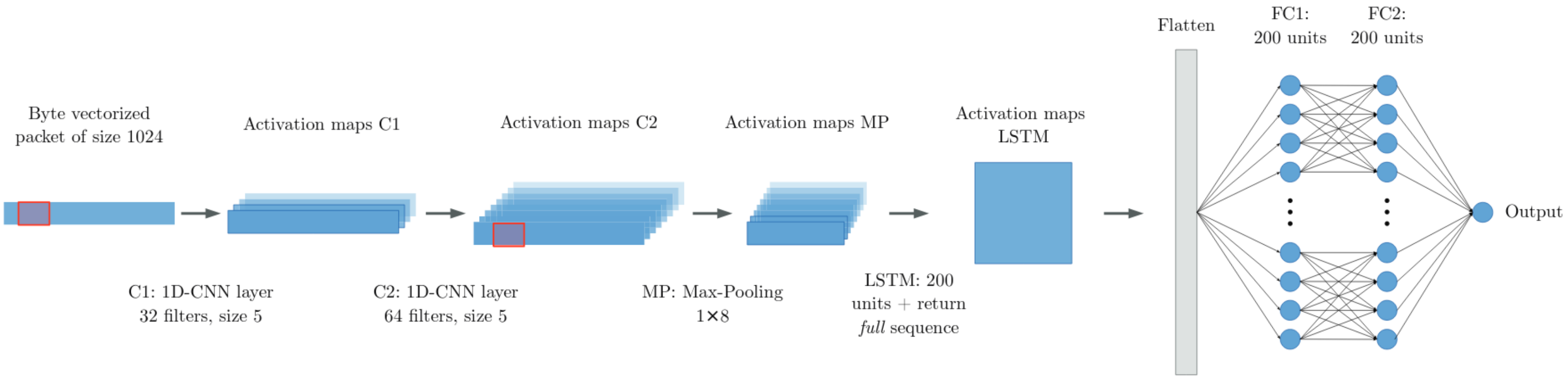
- *N-Gram* based feature extraction on application layer payloads
- Prototypical representations specific to individual types of messages
- Filtering rare features using a frequency threshold
- Cluster similarity based intrusion detection
- Takeaways:
  - **+** Evaluated on ICS related protocols
  - **+** Unsupervised feature extraction
  - **--** No sequential aspects are considered



Brandenburg University of Technology Cottbus - Senftenberg

[3] Deep in the Dark - Deep Learning-based Malware Traffic Detection without Expert Knowledge *(IEEE SPW, 2019)*

- Feature extraction on first *N* bytes of every packet / flow
- MAC & IP addresses are sanitized
- Softmax based classification
- Takeaways:
  - -- End-to-end model
  - -- Supervised learning
  - -- Evaluation on ICS unrelated data
  - \+ Spatial–temporal representation learning on raw traffic



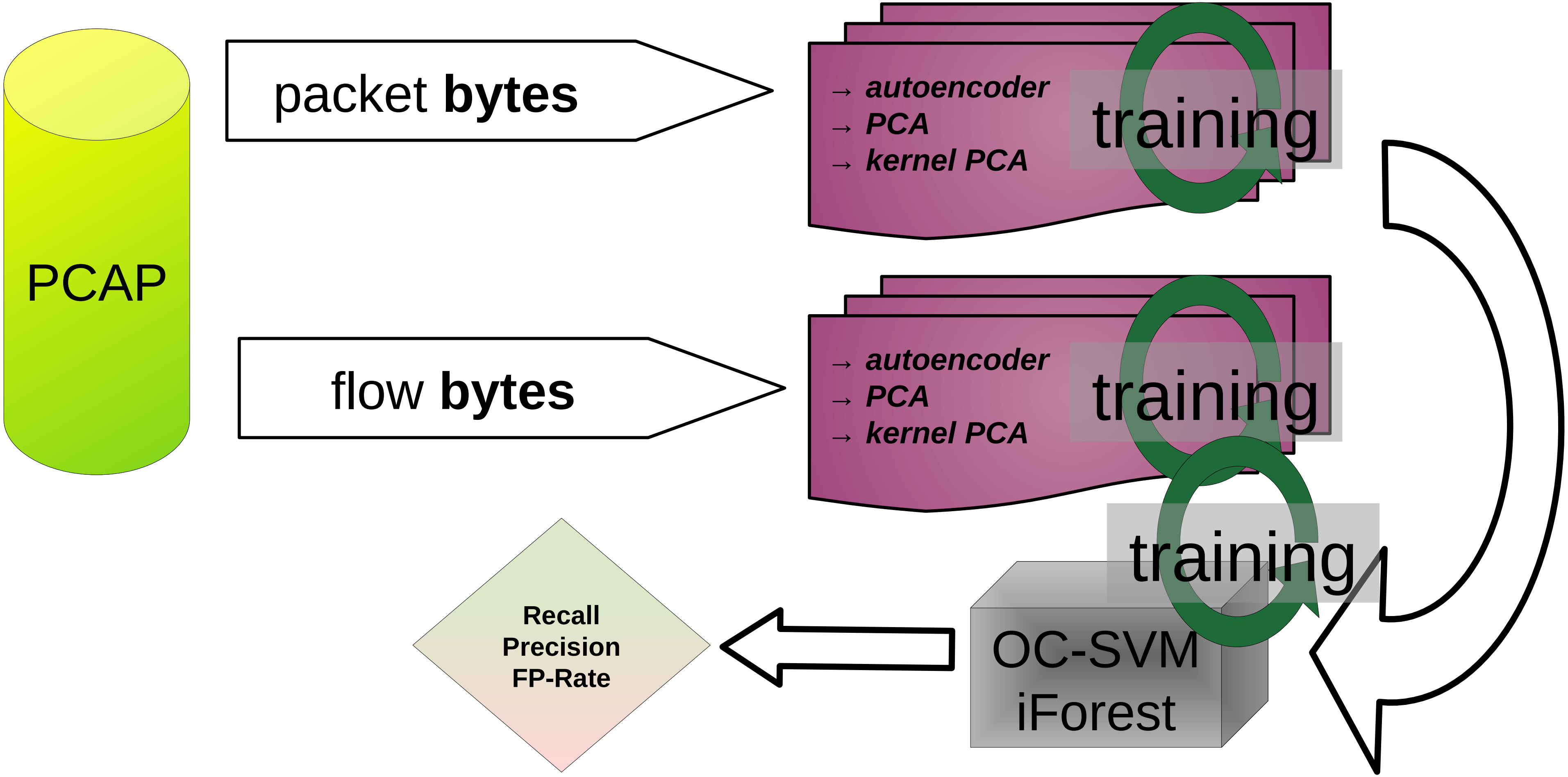
- **Can AD based on unsupervised feature extraction compete with manual feature extraction?**
  - *Point* anomalies
  - *Contextual* anomalies

- **What kind of byte representation will yield the best results?**
  - *PCAP bytes*
  - *Packet bytes*
  - *Flow bytes*

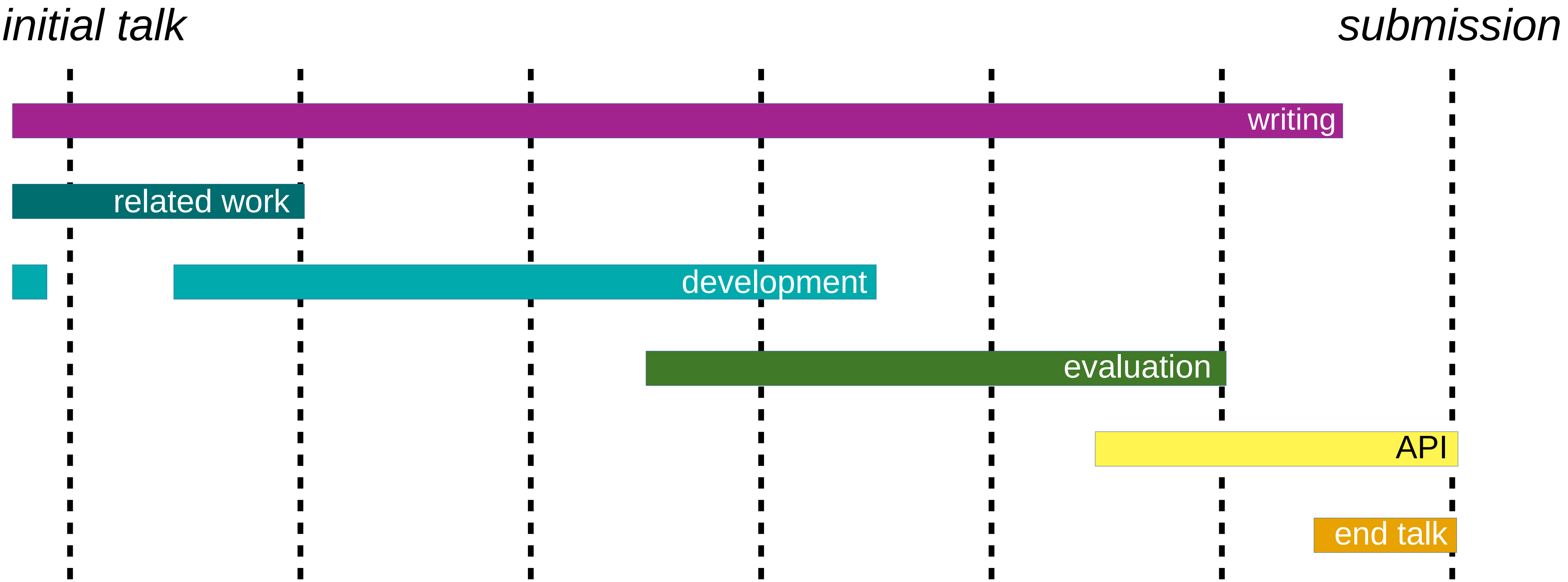
- **Is the approach *fast* enough ?**
  - Intrusion detection is a real time problem

- **Is the approach capable of extracting relevant features when analyzing variable traffic?**

Brandenburg
University of Technology
Cottbus - Senftenberg



Brandenburg University of Technology Cottbus - Senftenberg



Thanks for the attention!
Questions?

Sources

[1] Chandola, et al. *Anomaly detection: A survey* **ACM computing surveys, 2009**.

[2] *Wressnegger C., Ansgar K. & Konrad R.*

Zoe: Content-based anomaly detection for industrial control systems.

IEEE/IFIP: International Conference on Dependable Systems and Networks (**DSN**). **IEEE, 2018.**

[3] Marín G., Casas P. & Capdehourat, G.

Deep in the Dark-Deep Learning-Based Malware Traffic Detection Without Expert Knowledge.

IEEE: Security and Privacy Workshops (**SPW**). **IEEE, 2019.**

[i] https://i2.wp.com/thedatascientist.com/wp-content/uploads/2019/02/anomaly_detection.png

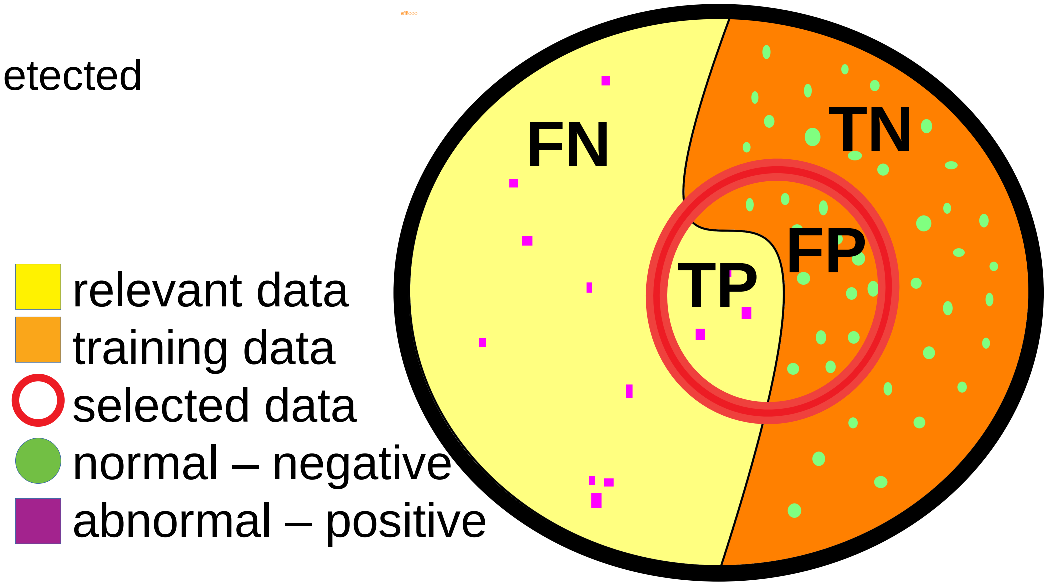
[ii] https://miro.medium.com/max/1000/0*sQzmiOf8Yb_18HX1.png

[iii] Saul, Lawrence K., et al.

"Spectral methods for dimensionality reduction" **Semi-supervised learning 3**, **2006**.

Appendix | *Evaluation*

False Negative (**FN**) – abnormal data that was not detected
True Positive (**TP**) – detected abnormal data
True Negative (**TN**) – correctly ignored normal data
False Positive (**FP**) – normal data wrongly detected



metrics are derived from the confusion matrix

- Recall $\mathrm{RE}=\frac{FP}{FN+TP}$
- Precision $PR=\frac{FP}{FP+TP}$
- FP-rate $FPR=\frac{FP}{N+P}$
- F_1-Score $F1=2*(\frac{\mathrm{RE}*PR}{\mathrm{RE}+PR})$

Problem: one needs ground truth information for evaluation!

Concept

- Learns the identity function **ID(x) = x**
- *Encoder* network compresses representation → *code*
- *Decoder* network restores sample from *code*
- *Information bottleneck* forces generalization

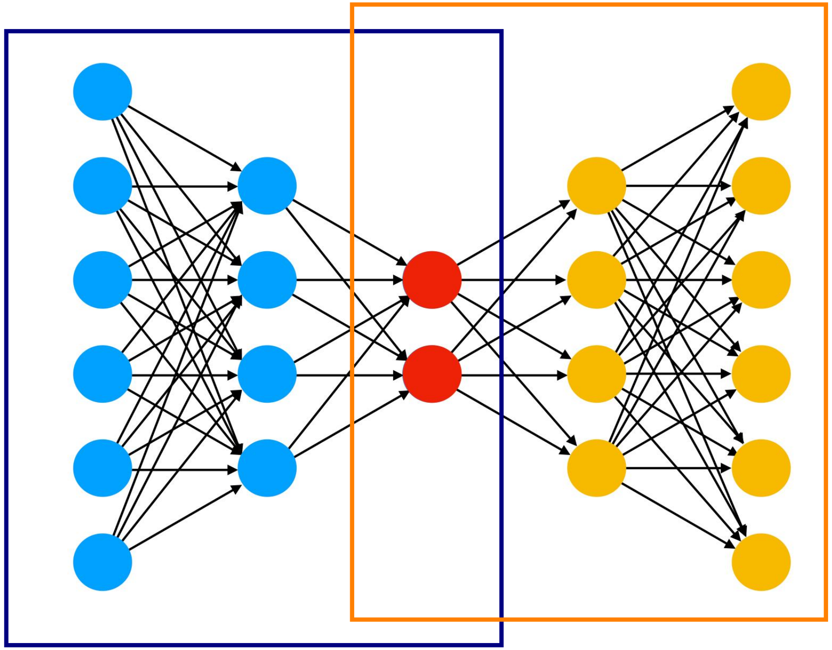
Trivia

- Learn unsupervised
- Publicized in late 60's ***[7]***
- Reconstruction *not* loss less
- Used in many different architectures (???)
- Neural networks are designed to simulate memory

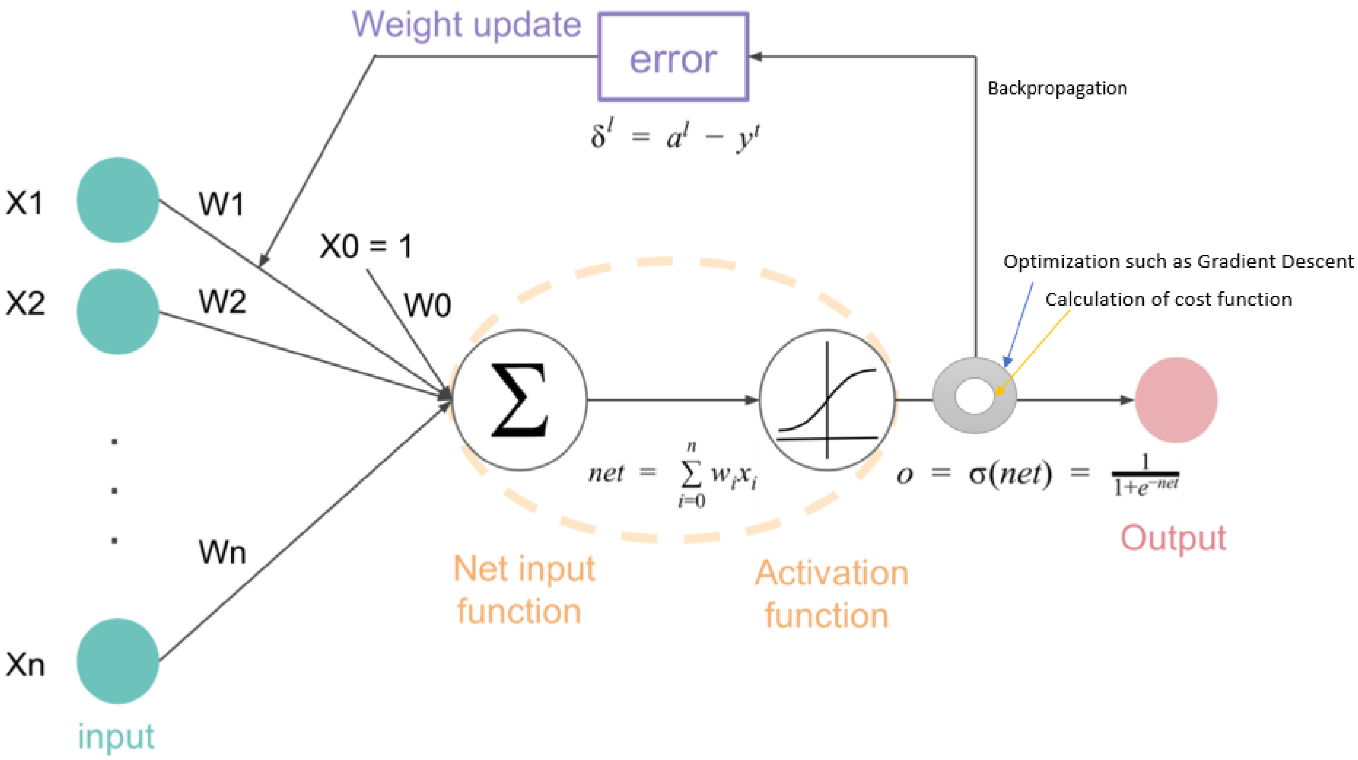
Pro's & Con's

+ Out of sample model
+ Gets better with data
+ Works unsupervised

-- Blackbox by design



Appendix | *Artificial Neuron*

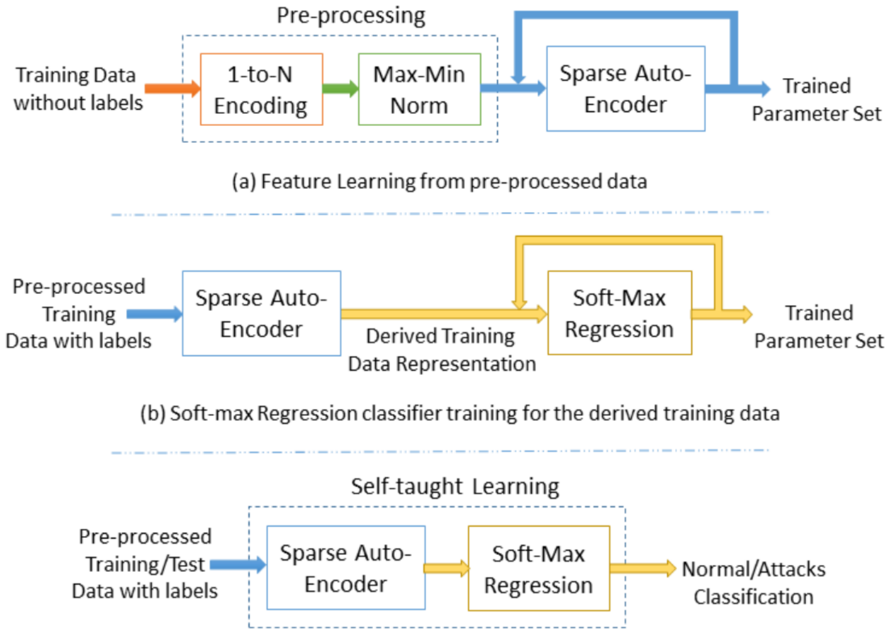


[https://i.stack.imgur.com/7Ui1C.png]

Appendix | **Related Work**

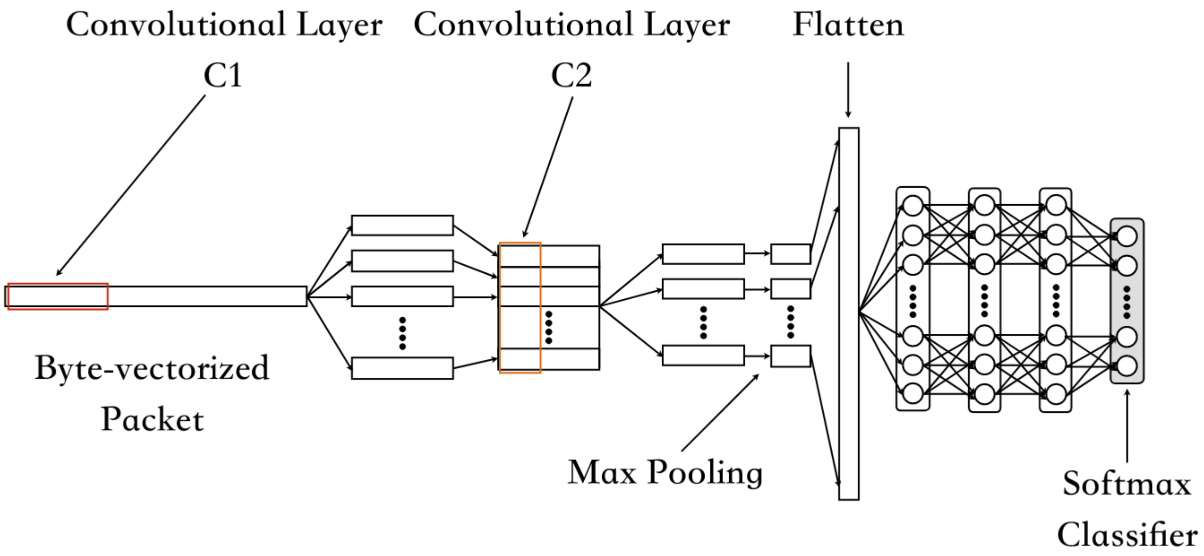
A Deep Learning Approach for Network Intrusion Detection System *(ACM BIONETICS, 2016)*

- *NSL-KDD* dataset (41 features)
- Autoencoder for unsupervised feature learning + Soft-max regression for for classification



Deep packet: a novel approach for encrypted traffic classification using deep learning *(Soft Computing, 2020)*

- *Traffic characterization* (FTP, P2P, ...) and *application identification* (BitTorrent, Skype, …)
- Distinguishes between VPN and nonVPN traffic, but fails to classify *tor* traffic
- Comparison between different supervised architectures (SAE and CNN)
- UNB ISCX dataset
- Do not regard any temporal phenomenon

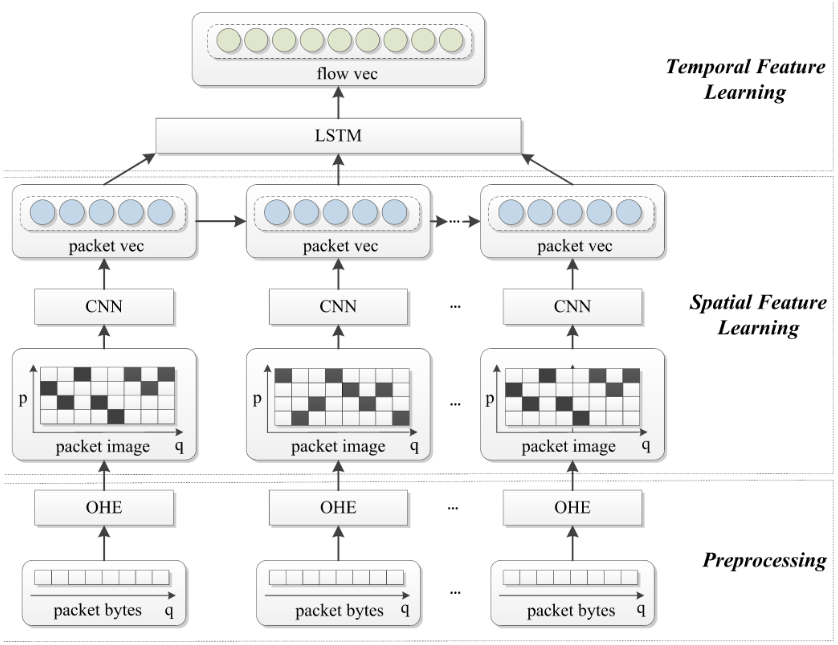


Appendix | **Related Work**

Brandenburg University of Technology Cottbus - Senftenberg

HAST-IDS: Learning Hierarchical Spatial-Temporal Features Using Deep Neural Networks to Improve Intrusion Detection *(IEEE ACCESS, 2017)*

- *DARPA1998* and *ISCX2012* for evaluation
- Bytes are transformed via a one-hot-encoding
- Soft-max for classification



Appendix | **Related Work**

[2] Malware Traffic Classification Using Convolutional Neural Network for Representation Learning *(IEEE ICOIN*, *2017)*

- *USTC-TFC2016* data set
- *Spatial* feature extraction (*LeNet-5)* + Soft-Max regression classifier
- Bi-directioal packet representation with all layers yields best results