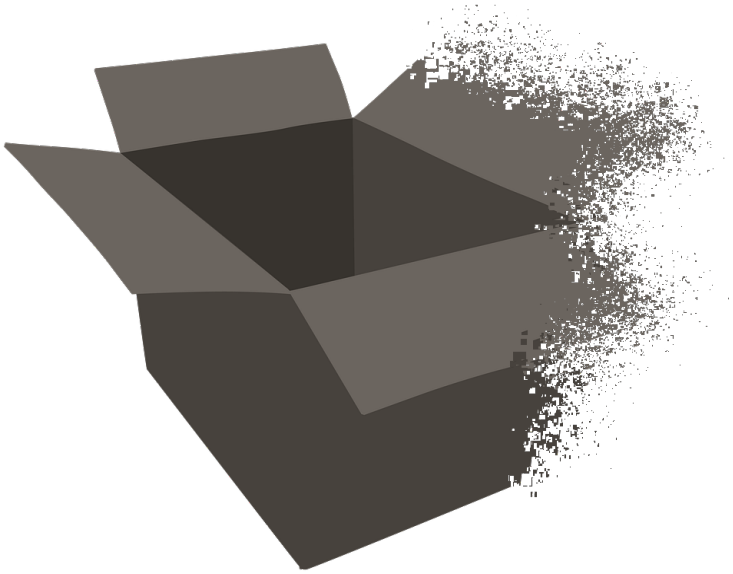


Faculty 1: Chair of IT-Security | Master Thesis

Representation Learning for Content-Sensitive Anomaly Detection in Industrial Networks

B. Sc. Fabian Kopp

Supervisors:

M. Sc. Franka Schuster

Prof. Dr.-Ing. Andriy Panchenko

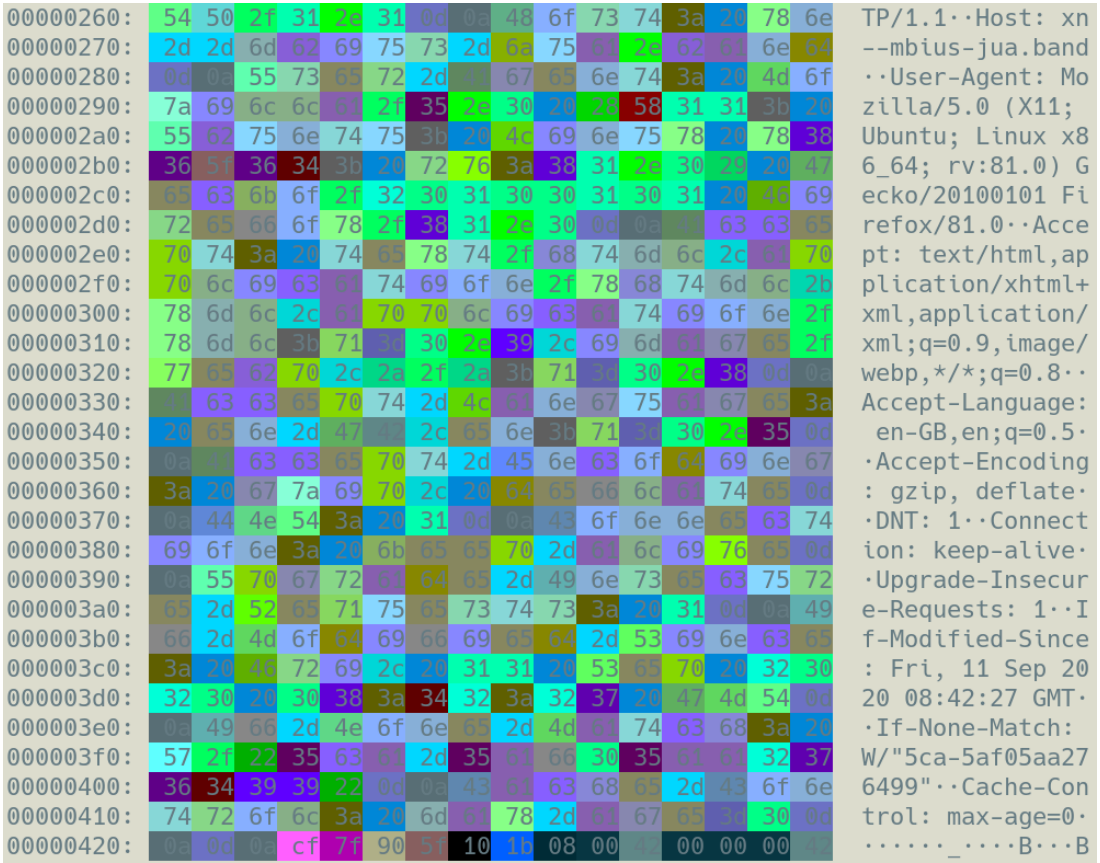
Brandenburg University of Technology Cottbus - Senftenberg

Network Data

- Series of *packets*
- Weak order of packets
- Packets consist of *layers*
- Different types of protocols
- Layers have *header* and *payload* fields

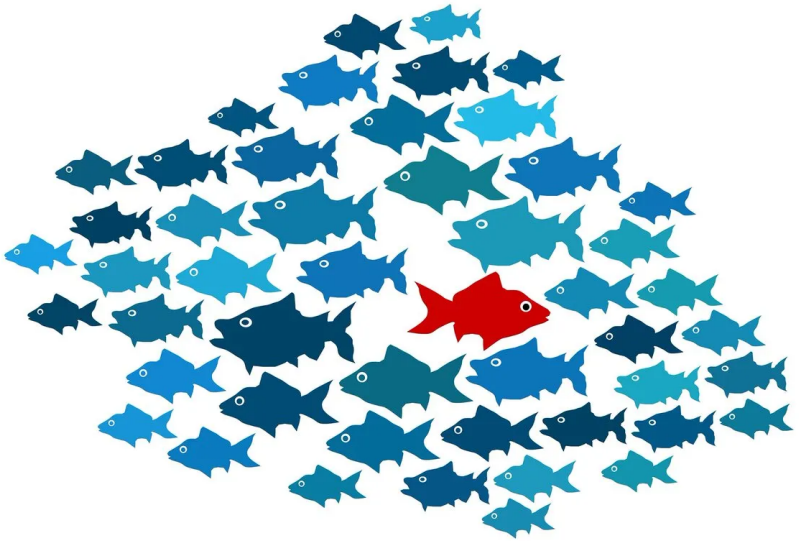
Industrial Control Systems (*ICS*)

- Monitoring & Control functions
- Fixed network topology
- Proprietary Infrastructure
- Cycle information exchange



Anomaly detection (*AD*)

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



Types of anomalies from *[1]*

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

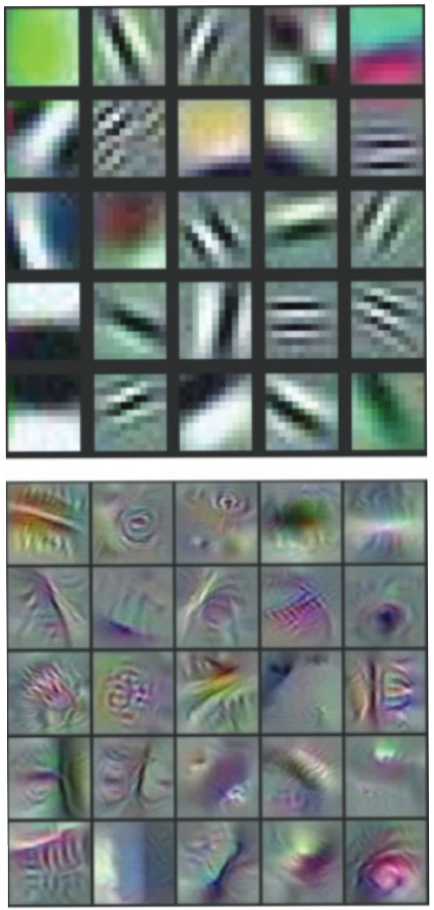
[1] Chandola et al: *Anomaly Detection : A Survey* | ACM 2009

Type of feature extraction

Capture essence of data while removing noise

Embedding of feature vectors

- Linear methods
  - Principal component analysis (PCA)
- Non-Linear methods
  - Kernel-PCA
  - Autoencoders (AE)
  - t-SNE



Automatic fish species classification in underwater videos: Exploiting pretrained deep neural network models to compensate for limited labelled data

Represent network captures (PCAP) as *pictures*

- 32x32=1024 bytes per fragment
- Different fragment pre-processing heuristics

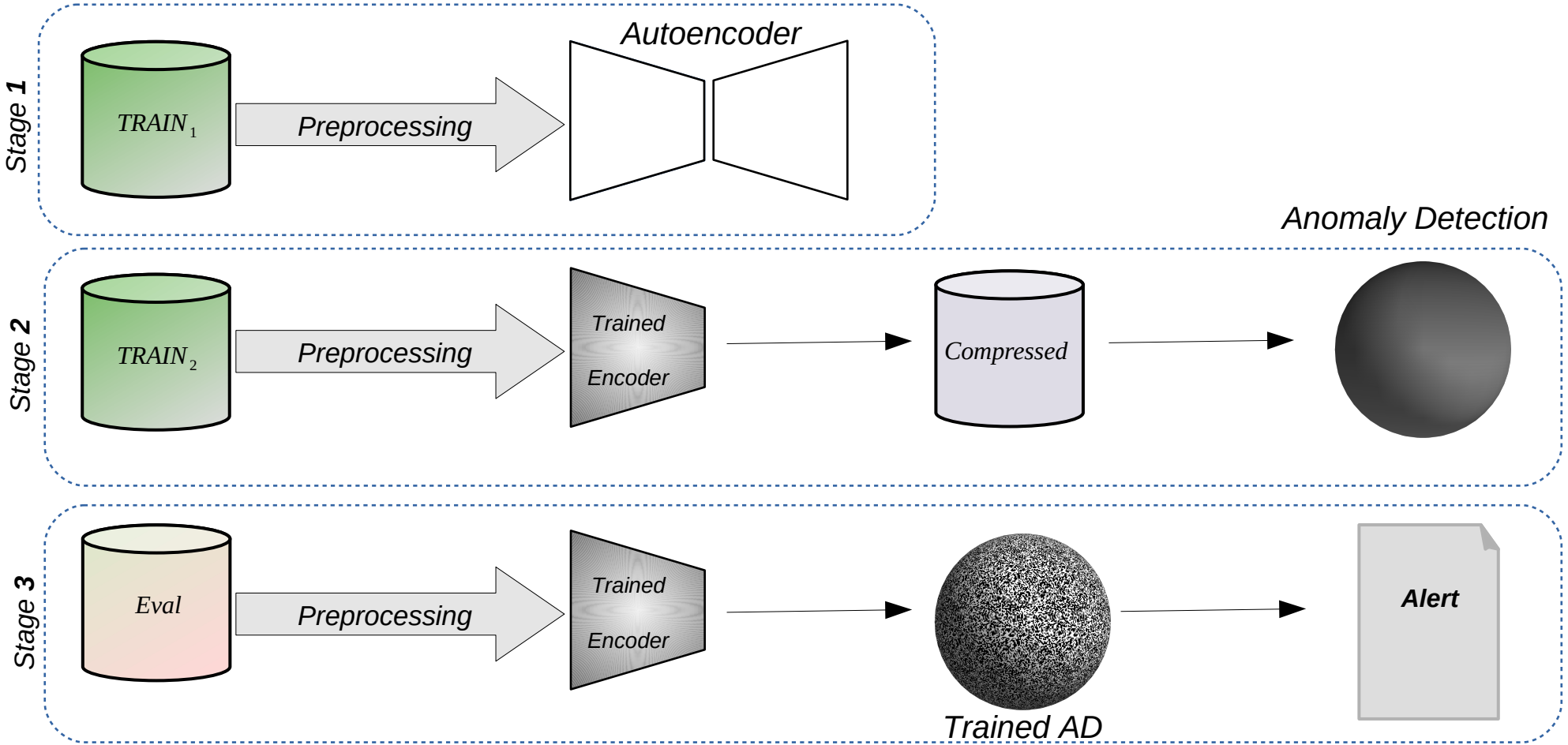
Unsupervised representation learning via autoencoder

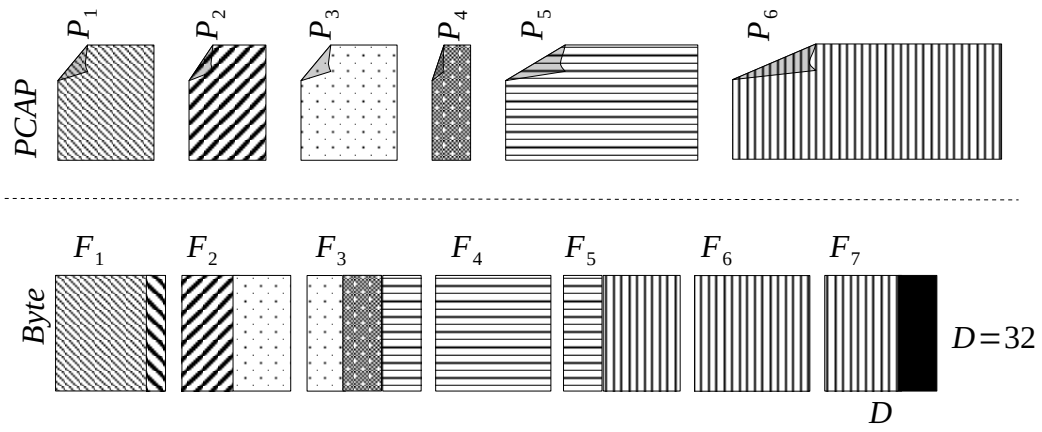
- ConvLSTM-Layer via: *Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting [arXiv:1506.04214]*
- Spatial-Temporal feature extraction
- 4 layers per en/de-coder
- Bottleneck: $4 \times 4 \times 4 = 64 > \sqrt{1024}$

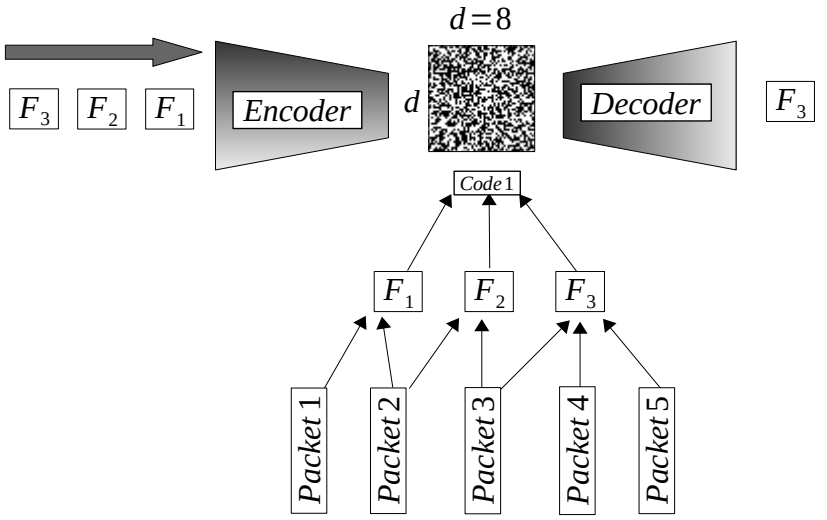
Use learned representation to perform classic shallow anomaly detection

- One-Class Support Vector Machine (OCSVM)
- Isolation Forest (IF)
- Local Outlier Factor (LOF)

- **Can shallow AD work with unsupervised feature extraction ?**
  - *Point* anomalies
  - *Contextual* anomalies
- **What kind of raw representation yields best results?**
  - *Byte orientation*
  - *Packet orientation*
  - *Flow orientation*
- **Can anomalies automatically be explained ?**
  - *Close the semantic gap*

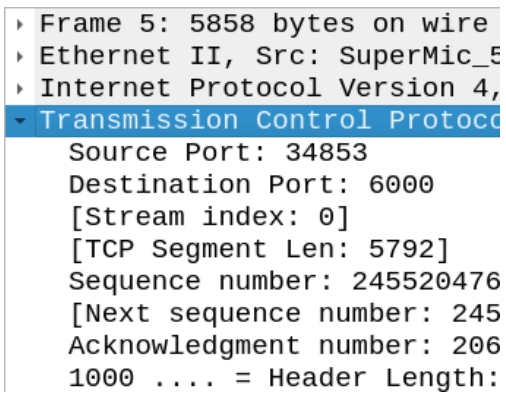

Stage 1
TRAIN1
Preprocessing
Autoencoder
Anomaly Detection
Stage 2
TRAIN2
Preprocessing
Trained
Encoder
Compressed
Stage 3
Eval
Preprocessing
Trained
Encoder
Trained AD
Alert


PCAP
P_1
P_2
P_3
P_4
P_5
P_6
Byte
F_1
F_2
F_3
F_4
F_5
F_6
F_7
$D=32$
D

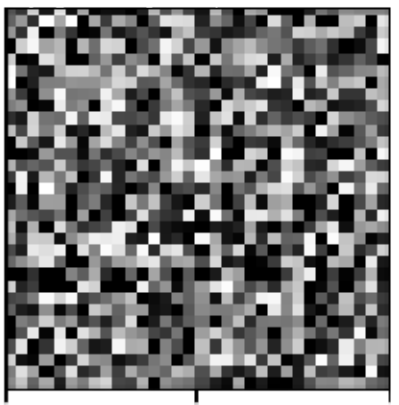

$d=8$
F_3
F_2
F_1
Encoder
d
Decoder
F_3
Code 1
F_1
F_2
F_3
Packet 1
Packet 2
Packet 3
Packet 4
Packet 5

Available Information for Alerts

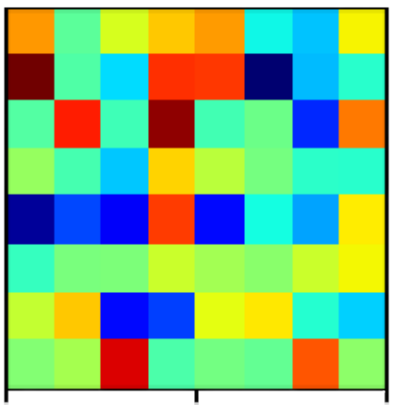
- Anomaly Score (0,1]
- Input Packets
- Latent representation
- Input fragment representation
- Model loss (mean squared error)



Input Packet



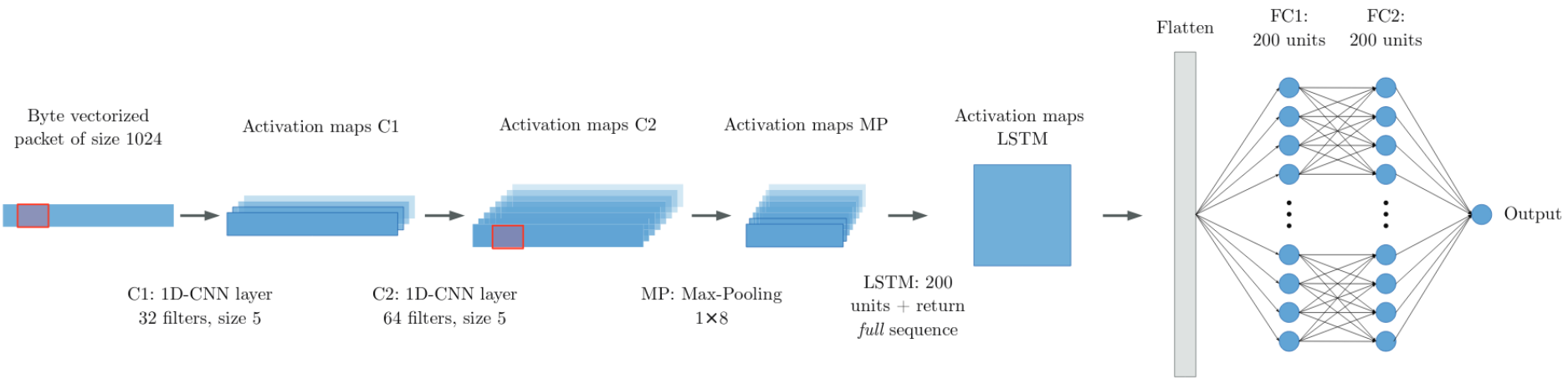
Input Fragment



Latent Representation

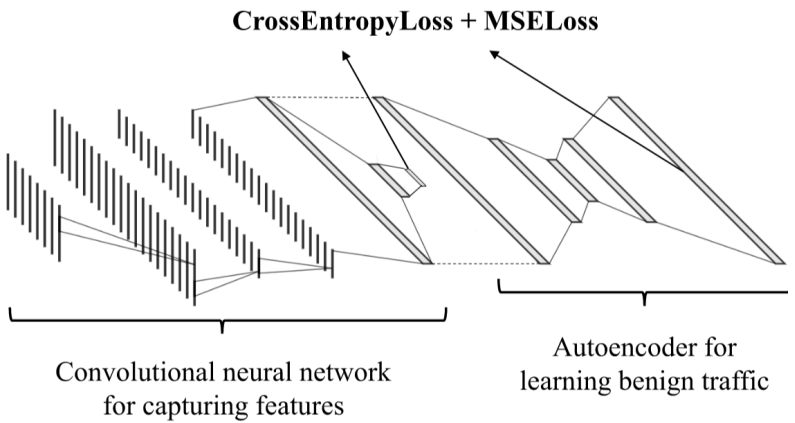
DeepMAL - Deep Learning Models for Malware Traffic Detection and Classification
(iDSC2020, 2020)

- Feature extraction on first *1000* bytes of every packet / flow
- Spatial–Temporal representation learning on raw traffic
- Supervised softmax-based classification
- Comparison to shallow methods & handcrafted features extraction



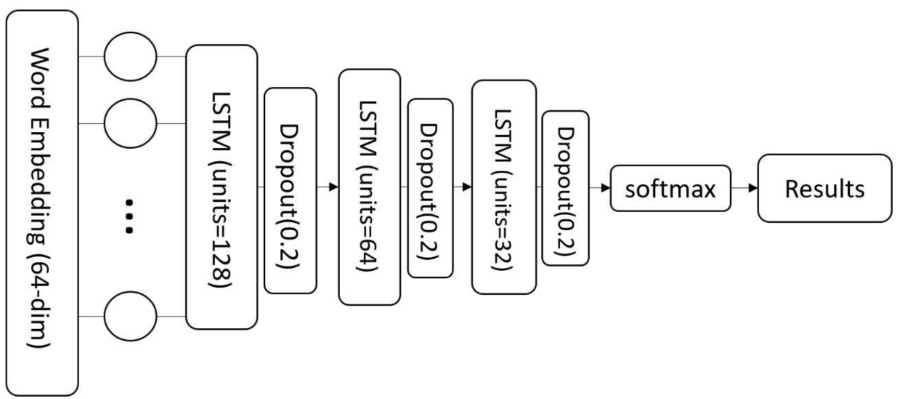
An Unsupervised Deep Learning Model for Early Network Traffic Anomaly Detection
(IEEE Access, 2020)

- *1*D-CNN for representation learning
- Loss based anomaly detection
- Sampling technique to handle large network throughput
- First 80 bytes per sample is used for detection
- Evaluation on DoS related data sets



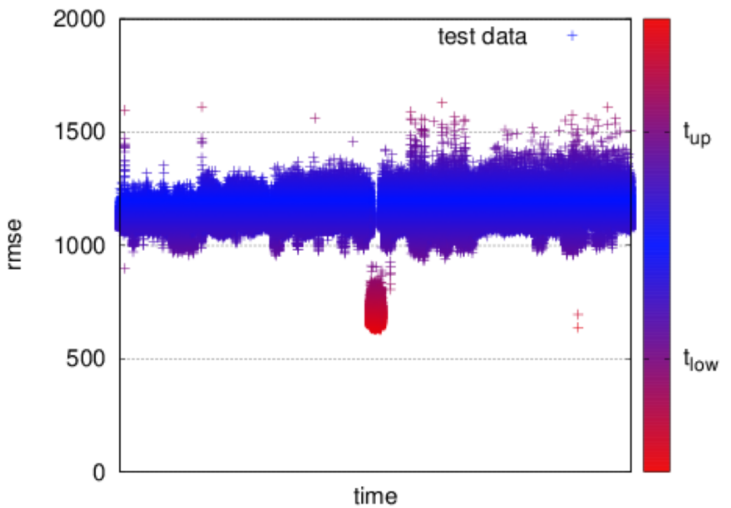
An LSTM-Based Deep Learning Approach for Classifying Malicious Traffic at the Packet Level
(Applied Sciences, 2019)

- Extract protocol features via unsupervised word embeddings
- Packet orientation approach
- Focus on protocol headers
- LSTM encoding & softmax-based classification



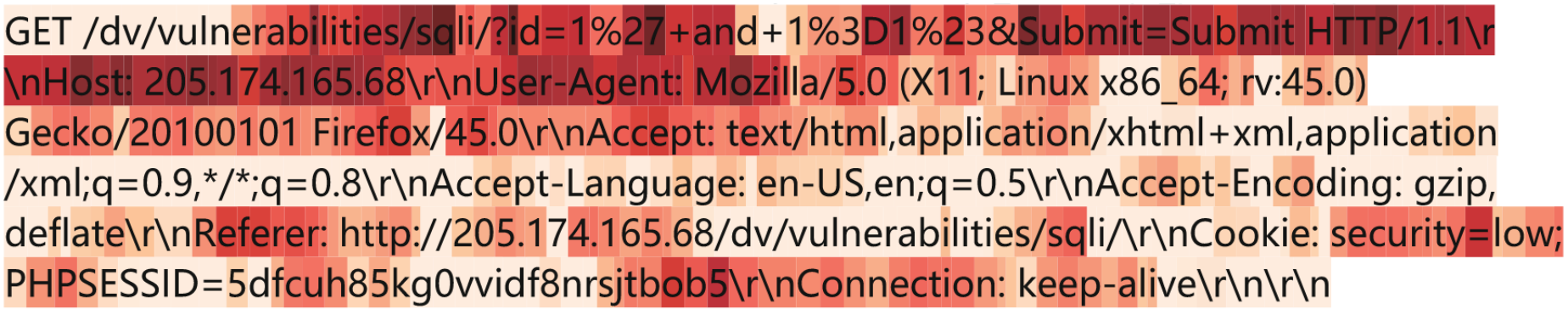
High-Performance Unsupervised Anomaly Detection for Cyber-Physical System Networks
(CPS-SPC, 2018)

- Unsupervised feature learning
- First 1000 bytes used for detection
- 3 layer linear denosing autoencoder
- No direct temporal learning
- Loss based anomaly detection
- Evaluation on SWaT data set



Attentional Payload Anomaly Detector for Web Applications
(ICONIP, 2018)

- Anomaly detection based on raw bytes sequences
- TCP payloads are basis for detection
- *Attention* based RNN model
- Supervised training
- Heuristic to visualize the models' interpretation



Brandenburg University of Technology Cottbus - Senftenberg

➔ Semantic gap vs. SOTA

- High f1-Score is important, but there are other goals!
- NLP better suited to close *semantic gap*?

➔ Transparency & Interpretability

- Feature Visualization [distill.pub/2017/feature-visualization]
  - *Idea*: visualize a prototype of a normal e.g: TCP Header?
  - Does not help semantic gap
- Layer-wise Relevance Propagation (LRP) [heatmapping.org]
  - *Goal*: Heatmaps of show anomalous bytes!
  - Could not integrate TU-Berlin repo (RNN-autoencoder no softmax classifier)

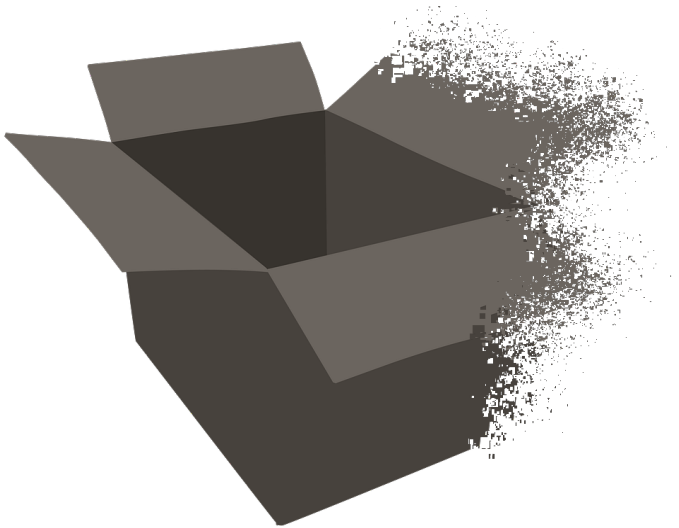
➔ Experience with papers ?

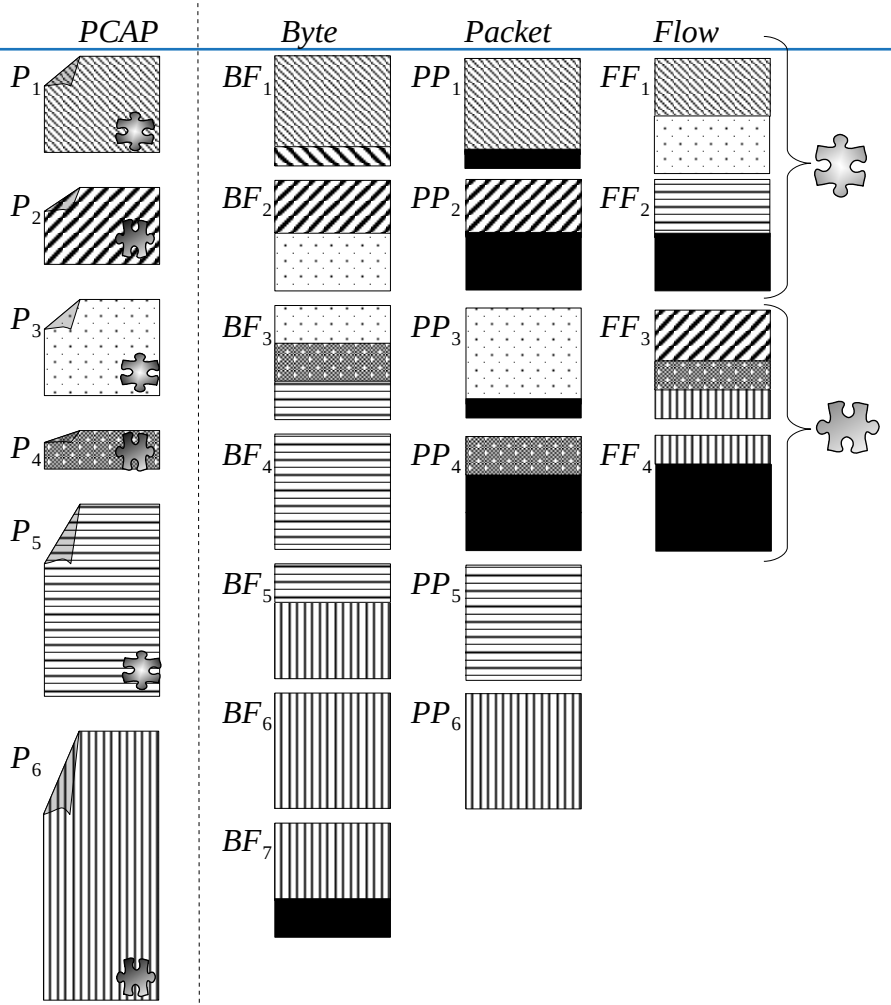
- Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting
  - Spatial-Temporal Neurons
- Explainable Deep One-Class Classification (Klaus-Robert Müller et. la)
  - Differentiable SVM

➔ The quest for a new benchmark data set...

- Data set for fully unsupervised anomaly detection on raw data
- How to turn SWaT into the new KDD?
- Respect the semantic gap!

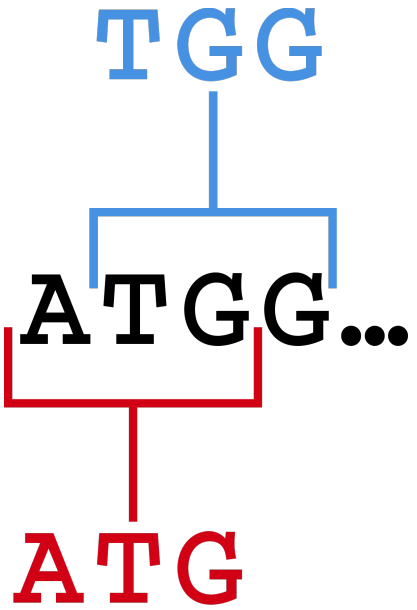
Thanks for the attention!
Questions?




PCAP
Byte
Packet
Flow
P_1
P_2
P_3
P_4
P_5
P_6
BF_1
BF_2
BF_3
BF_4
BF_5
BF_6
BF_7
PP_1
PP_2
PP_3
PP_4
PP_5
PP_6
FF_1
FF_2
FF_3
FF_4

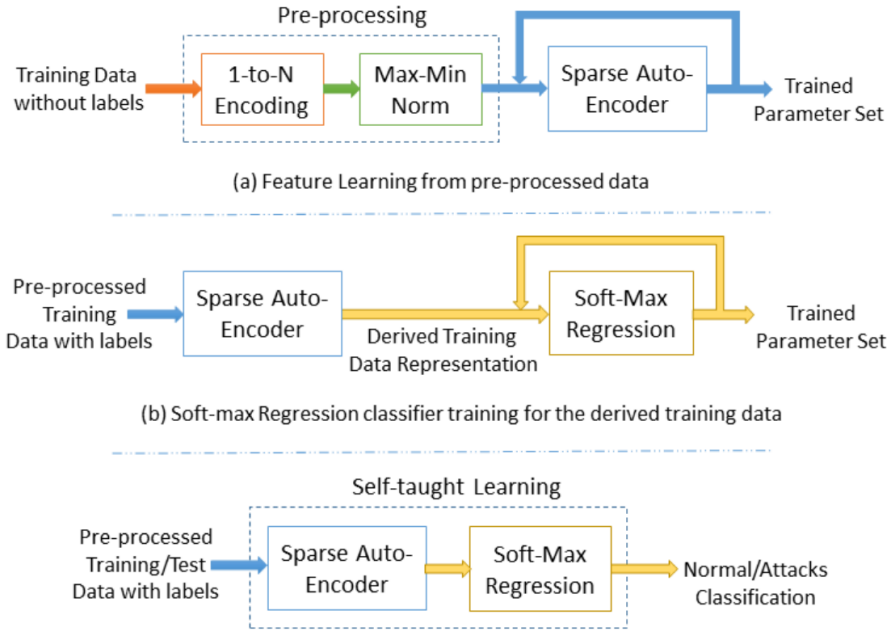
[2] ZOE: Content-based Anomaly Detection for Industrial Control Systems
(IEEE/IFIP DSN, 2018)

- *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
  - **--** Packet sequence not regarded



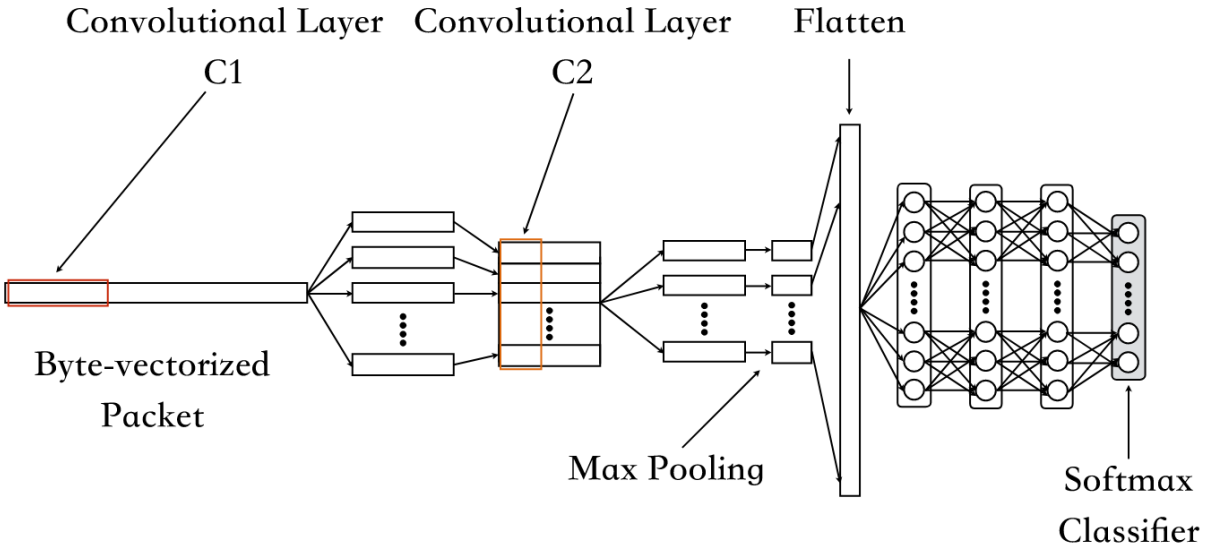
A Deep Learning Approach for Network Intrusion Detection System *(ACM BIONETICS, 2016)*

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



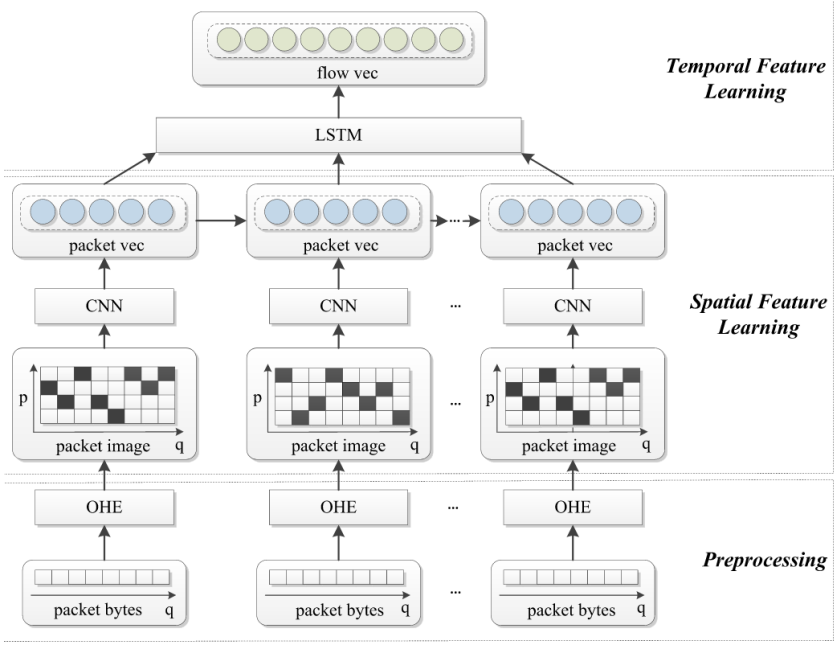
Deep packet: a novel approach for encrypted traffic classification using deep learning *(Soft Computing, 2020)*

- *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



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



[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