

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

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

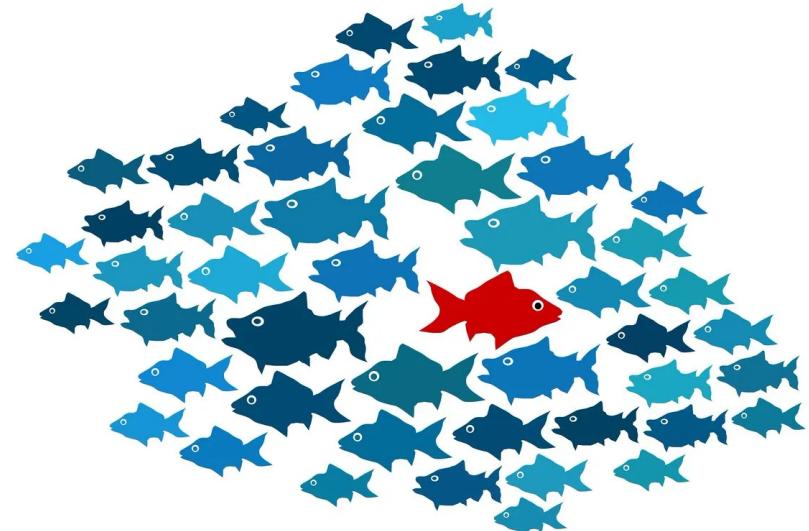
| | | | | | | | | | | | | | | | | | | |
|-----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|--|-------------------|
| 00000260: | 54 | 50 | 2f | 31 | 2e | 31 | 0d | 0e | 48 | 6f | 73 | 74 | 3a | 20 | 78 | 6e | TP/1.1 · Host: xn | |
| 00000270: | 2d | 2d | 6d | 62 | 69 | 75 | 73 | 2d | 6a | 75 | 61 | 2e | 62 | 61 | 6e | 64 | --mbius-jua.band | |
| 00000280: | 0d | 07 | 55 | 73 | 65 | 72 | 2d | 41 | 67 | 65 | 6e | 74 | 3a | 20 | 4d | 6f | · · User-Agent: Mo | |
| 00000290: | 7a | 69 | 6c | 6c | 61 | 21 | 35 | 2e | 30 | 20 | 28 | 58 | 31 | 31 | 3b | 20 | zilla/5.0 (X11; | |
| 000002a0: | 55 | 62 | 75 | 6e | 74 | 75 | 3b | 20 | 4d | 69 | 6e | 75 | 78 | 20 | 78 | 38 | Ubuntu; Linux x8 | |
| 000002b0: | 36 | 51 | 36 | 34 | 3b | 20 | 72 | 76 | 3a | 38 | 31 | 2e | 30 | 29 | 20 | 47 | 6_64; rv:81.0) G | |
| 000002c0: | 65 | 63 | 6b | 6f | 21 | 32 | 30 | 31 | 30 | 30 | 31 | 30 | 31 | 20 | 46 | 69 | ecko/20100101 Fi | |
| 000002d0: | 72 | 65 | 66 | 6f | 78 | 2f | 38 | 31 | 2e | 30 | 0d | 0a | 41 | 63 | 63 | 65 | refox/81.0 · Acce | |
| 000002e0: | 70 | 74 | 3a | 20 | 74 | 65 | 78 | 74 | 2f | 68 | 74 | 6d | 6c | 2c | 61 | 70 | pt: text/html,ap | |
| 000002f0: | 70 | 6c | 69 | 63 | 61 | 74 | 69 | 6f | 6e | 21 | 78 | 68 | 74 | 6d | 6c | 2b | plication/xhtml+xml,application/xml+xml;q=0.9,image/webp,*/*;q=0.8 · | |
| 00000300: | 78 | 6d | 6c | 2c | 61 | 70 | 70 | 6c | 69 | 63 | 01 | 74 | 69 | 6f | 6e | 21 | Accept-Language: en-GB,en;q=0.5 · | |
| 00000310: | 78 | 6d | 6c | 3b | 71 | 50 | 30 | 2e | 39 | 2c | 69 | 6d | 61 | 67 | 65 | 20 | · · Accept-Encoding: gzip, deflate · | |
| 00000320: | 77 | 65 | 62 | 70 | 2c | 2a | 2f | 2a | 3b | 71 | 30 | 30 | 2e | 38 | 0d | 02 | · · DNT: 1 · · Connect ion: keep-alive · | |
| 00000330: | 4 | 63 | 63 | 65 | 70 | 74 | 2d | 4c | 61 | 6e | 67 | 75 | 61 | 67 | 65 | 3a | Upgrade-Insecur e-Requests: 1 · · I | |
| 00000340: | 20 | 65 | 6e | 2d | 47 | 42 | 2c | 65 | 6e | 3b | 71 | 3e | 30 | 2e | 35 | 0d | f-Modified-Since : Fri, 11 Sep 20 | |
| 00000350: | 0a | 4 | 63 | 63 | 63 | 65 | 70 | 74 | 2d | 45 | 6e | 63 | 6f | 64 | 69 | 6e | 67 | 20 08:42:27 GMT · |
| 00000360: | 3a | 20 | 67 | 7a | 69 | 70 | 2c | 20 | 64 | 65 | 66 | 6c | 61 | 74 | 65 | 0d | · · If-None-Match: W/"5ca-5af05aa27 | |
| 00000370: | 0a | 44 | 4e | 54 | 3a | 20 | 31 | 0d | 0a | 43 | 6f | 6e | 66 | 65 | 63 | 74 | 6499" · · Cache-Con | |
| 00000380: | 69 | 6f | 6e | 3a | 20 | 6b | 65 | 65 | 70 | 2d | 61 | 6c | 69 | 76 | 65 | 0d | trol: max-age=0 · | |
| 00000390: | 0a | 55 | 70 | 67 | 72 | 61 | 64 | 65 | 2d | 49 | 6e | 73 | 63 | 63 | 75 | 72 | _ · · B · · B | |
| 000003a0: | 65 | 2d | 52 | 65 | 71 | 75 | 65 | 73 | 74 | 73 | 3a | 20 | 31 | 0d | 0a | 49 | | |
| 000003b0: | 66 | 2d | 4d | 6f | 64 | 69 | 66 | 69 | 65 | 64 | 2d | 53 | 69 | 6e | 63 | 65 | | |
| 000003c0: | 3a | 20 | 46 | 72 | 69 | 2c | 20 | 31 | 31 | 20 | 53 | 65 | 70 | 20 | 32 | 30 | | |
| 000003d0: | 32 | 30 | 20 | 30 | 38 | 3a | 34 | 32 | 3a | 32 | 37 | 20 | 47 | 4d | 54 | 0d | | |
| 000003e0: | 0a | 49 | 60 | 2d | 4e | 6f | 6e | 65 | 2d | 4d | 61 | 74 | 63 | 68 | 3a | 20 | | |
| 000003f0: | 57 | 2f | 22 | 35 | 63 | 61 | 2d | 35 | 61 | 66 | 30 | 35 | 61 | 62 | 32 | 37 | | |
| 00000400: | 36 | 34 | 39 | 39 | 22 | 0d | 0a | 43 | 61 | 63 | 68 | 65 | 2d | 43 | 6f | 6e | | |
| 00000410: | 74 | 72 | 6f | 6c | 3a | 20 | 6d | 61 | 78 | 2d | 61 | 67 | 65 | 30 | 30 | 0d | | |
| 00000420: | 0a | 0d | 07 | cf | 7f | 90 | 5f | 10 | 1b | 08 | 00 | 41 | 00 | 00 | 00 | 00 | | |

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)



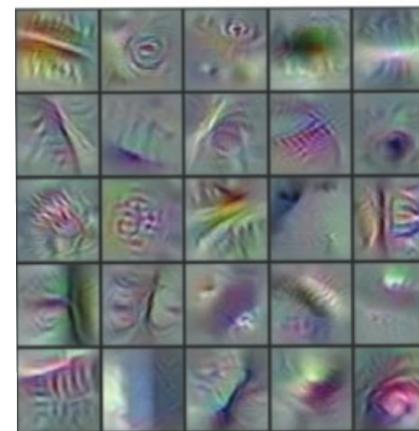
[1] Chandola et al: *Anomaly Detection : A Survey* | ACM 2009

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*

- $32 \times 32 = 1024$ bytes per fragment
- Different fragment pre-processing heuristics

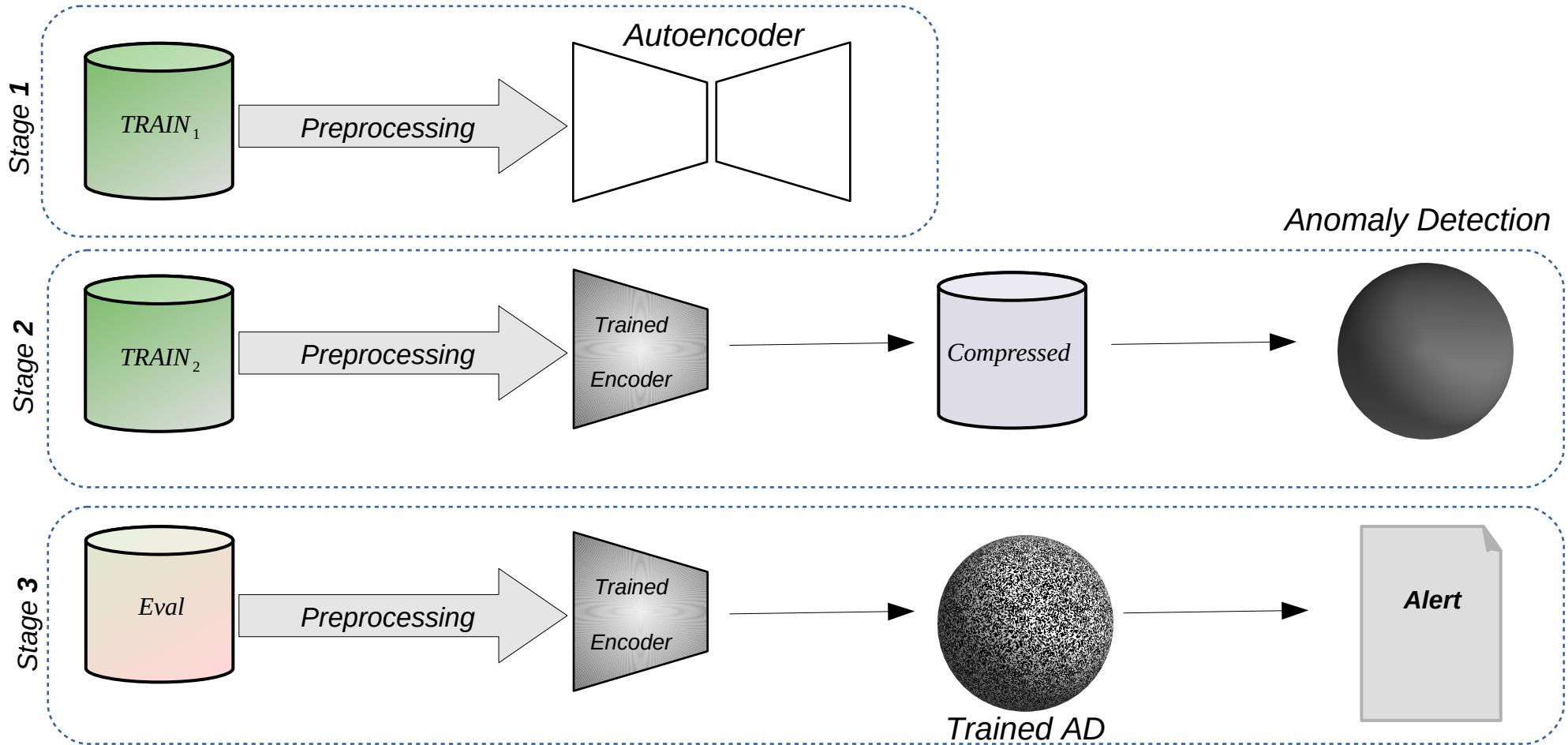
Unsupervised representation learning via autoencoder

- 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: $4 \times 4 \times 4 = 64 > \sqrt{1024}$

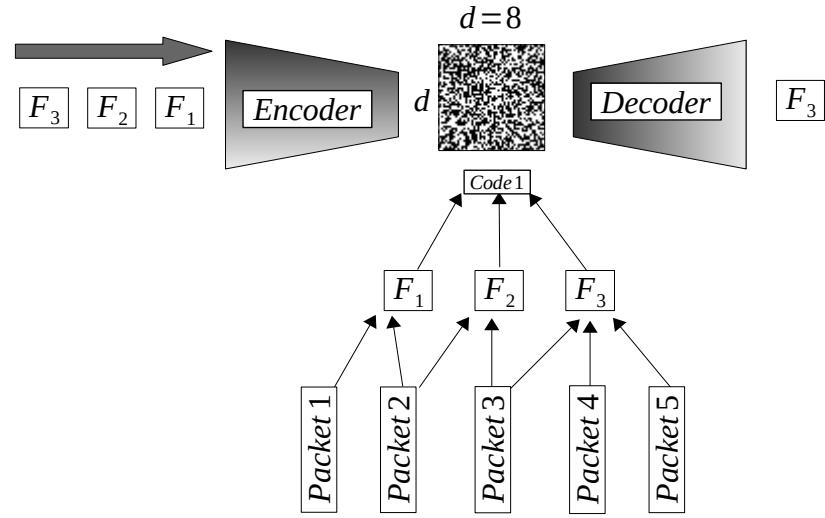
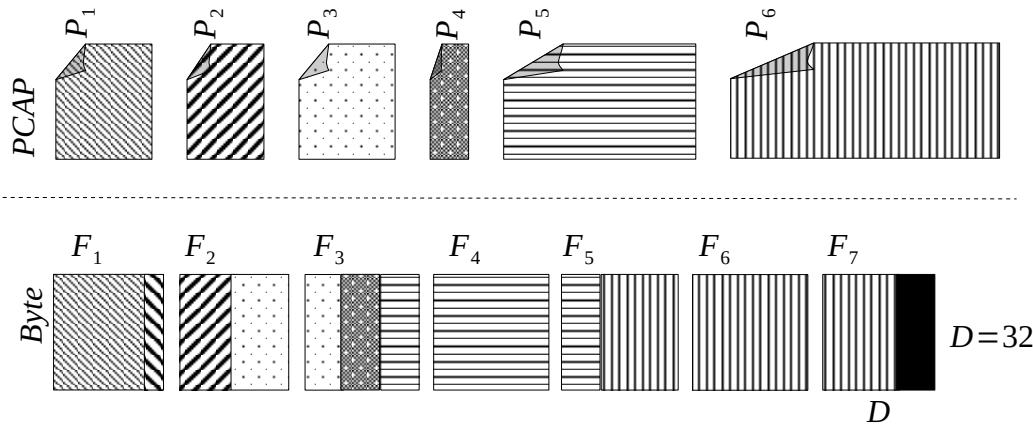
Use learned representation to perform classic shallow anomaly detection

- 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*



3 | Step 1: Representation Learning

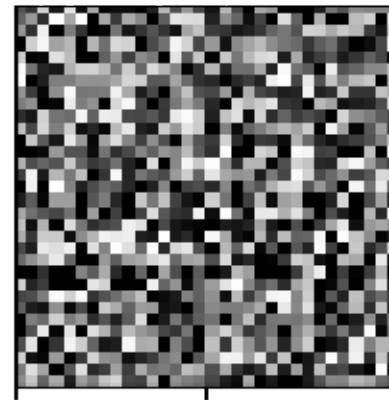


Available Information for Alerts

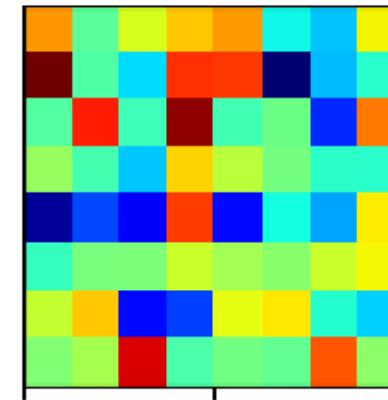
- Anomaly Score (0,1]
- Input Packets
- Latent representation
- Input fragment representation
- Model loss (mean squared error)

```
‣ Frame 5: 5858 bytes on wire
‣ Ethernet II, Src: SuperMic_5
‣ Internet Protocol Version 4,
- Transmission Control Protocol
    Source Port: 34853
    Destination Port: 6000
    [Stream index: 0]
    [TCP Segment Len: 5792]
    Sequence number: 245520476
    [Next sequence number: 245
    Acknowledgment number: 206
    1000 .... = Header Length:
```

Input Packet



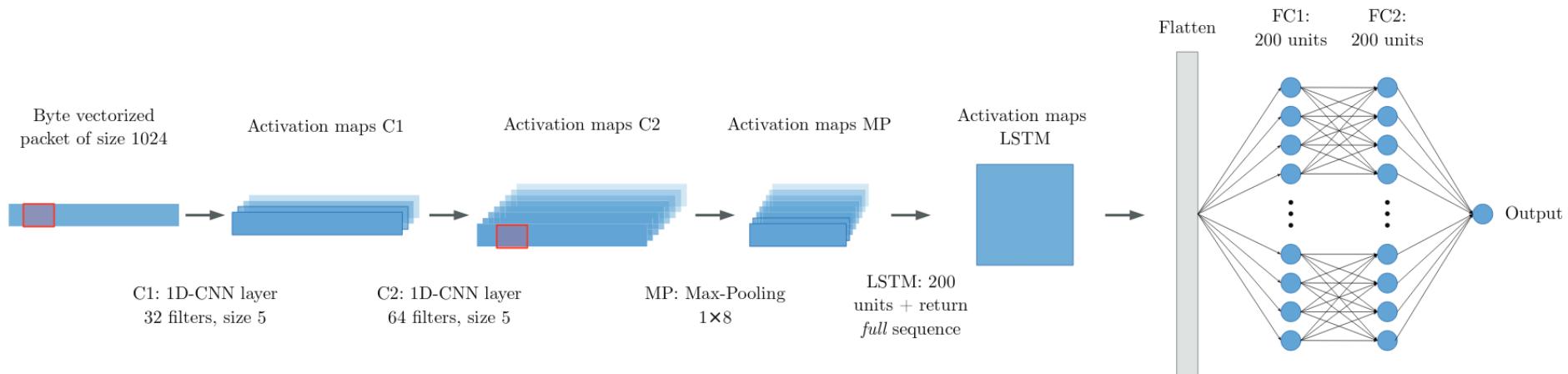
Input Fragment



Latent Representation

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

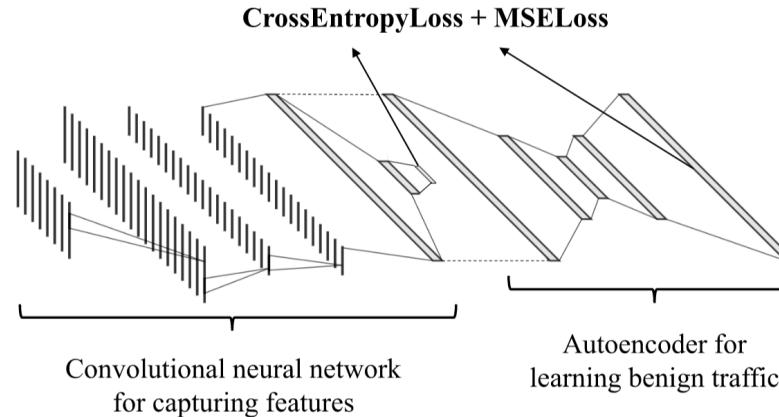
- Feature extraction on first 1000 bytes of every packet / flow
- Spatial-Temporal representation learning on raw traffic
- Supervised softmax-based classification
- Comparison to shallow methods & handcrafted features extraction



An Unsupervised Deep Learning Model for Early Network Traffic Anomaly Detection

(IEEE Access, 2020)

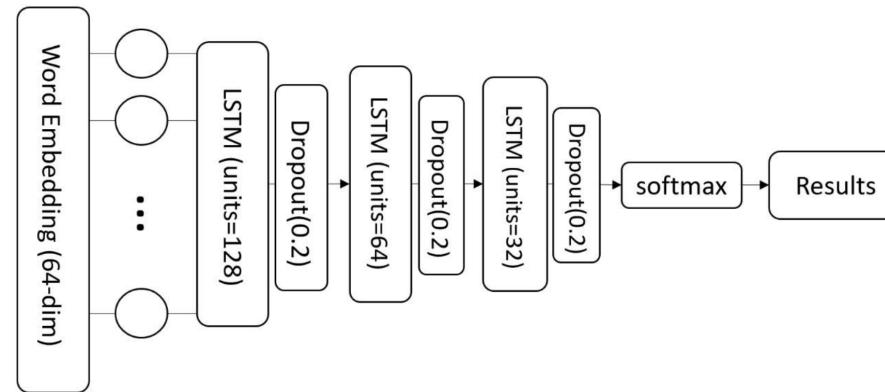
- 1D-CNN for representation learning
- Loss based anomaly detection
- Sampling technique to handle large network throughput
- First 80 bytes per sample is used for detection
- Evaluation on DoS related data sets



An LSTM-Based Deep Learning Approach for Classifying Malicious Traffic at the Packet Level

(*Applied Sciences*, 2019)

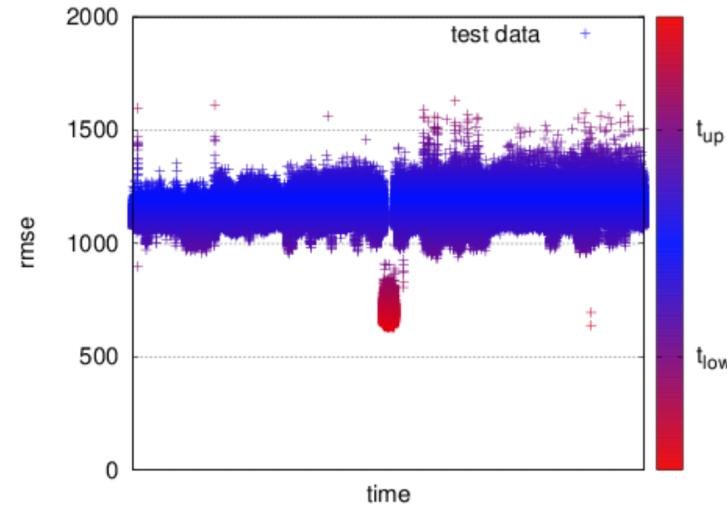
- Extract protocol features via unsupervised word embeddings
- Packet orientation approach
- Focus on protocol headers
- LSTM encoding & softmax-based classification



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

```
GET /dv/vulnerabilities/sqli/?id=1%27+and+1%3D1%23&Submit=Submit HTTP/1.1\r\nHost: 205.174.165.68\r\nUser-Agent: Mozilla/5.0 (X11; Linux x86_64; rv:45.0)\r\nGecko/20100101 Firefox/45.0\r\nAccept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8\r\nAccept-Language: en-US,en;q=0.5\r\nAccept-Encoding: gzip, deflate\r\nReferer: http://205.174.165.68/dv/vulnerabilities/sqli/\r\nCookie: security=low; PHPSESSID=5dfcuh85kg0vvvidf8nrsjtbob5\r\nConnection: keep-alive\r\n\r\n
```

→ 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)

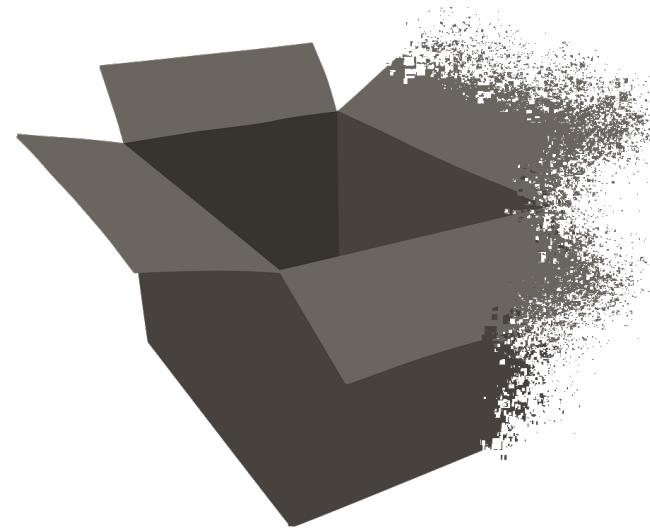
→ Experience with papers ?

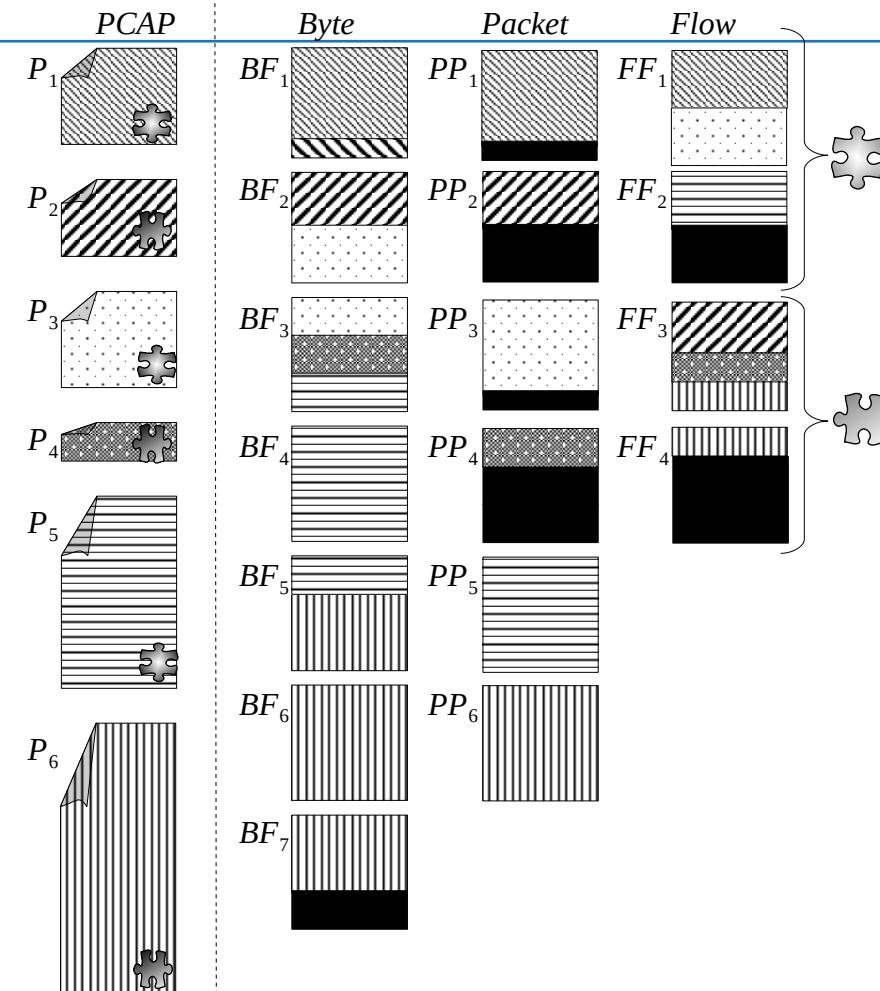
- Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting
 - Spatial-Temporal Neurons
- Explainable Deep One-Class Classification (Klaus-Robert Müller et. al)
 - Differentiable SVM

→ 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!

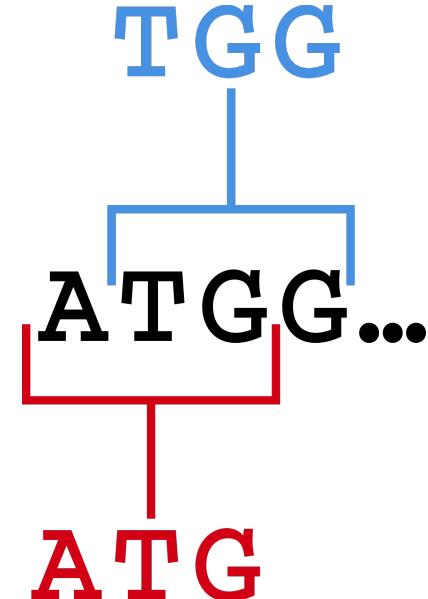
Thanks for the attention!
Questions?





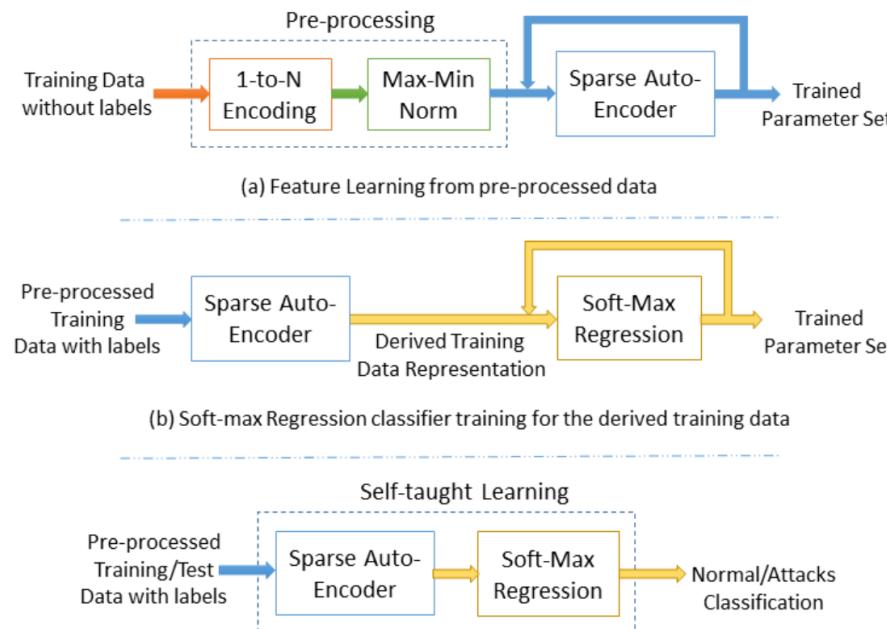
[2] ZOE: Content-based Anomaly Detection for Industrial Control Systems (IEEE/IFIP DSN, 2018)

- *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



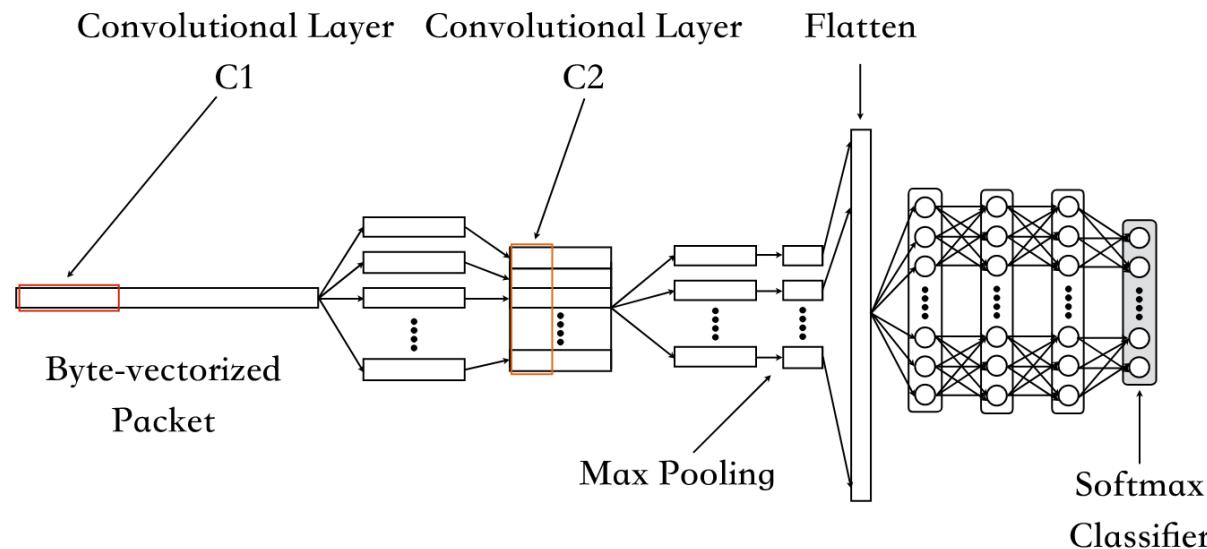
A Deep Learning Approach for Network Intrusion Detection System (*ACM BIONETICS, 2016*)

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



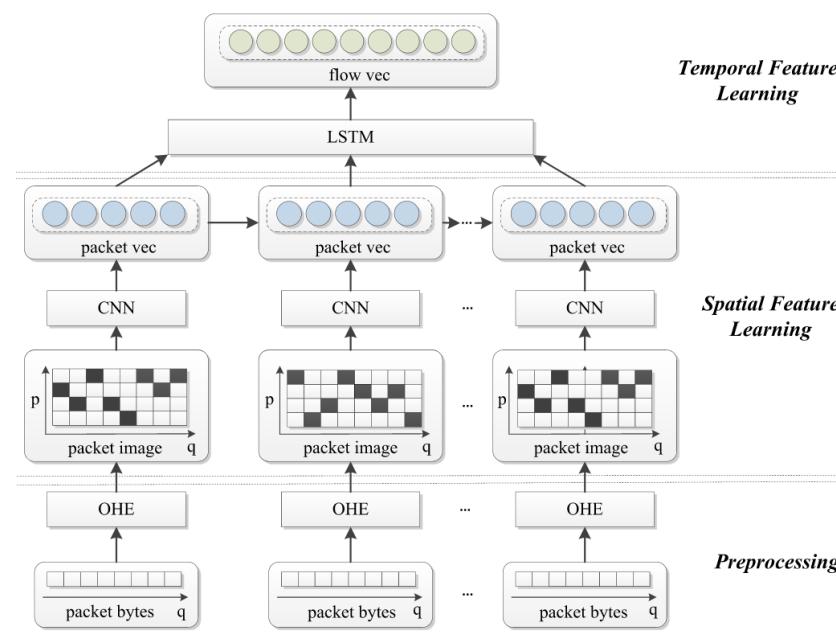
Deep packet: a novel approach for encrypted traffic classification using deep learning (*Soft Computing*, 2020)

- *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-directional packet representation with all layers yields best results

