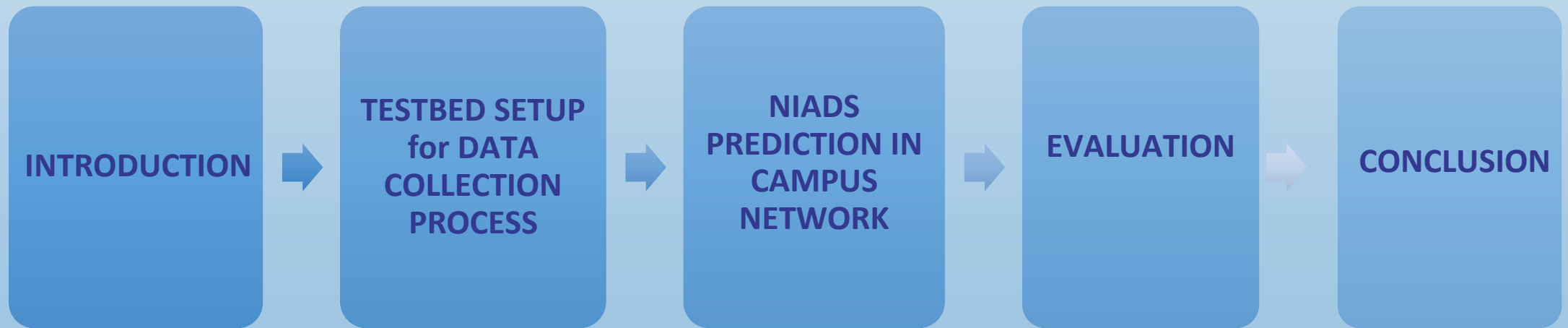


Machine Learning based Anomaly and Intrusion Detection to mitigate DoS and DDoS attacks in Private Campus Networks

****Sachinkumar B. Mallikarjun, *Mihiraj Dixit, and **Hans D. Schotten**

***Department of Computer Science, Saarland University, Saarbrücken
**WICON Chair, Department of Electrical and Computer Engineering,
University Of Kaiserslautern (RPTU), Kaiserslautern, Germany**





- **Motivation**

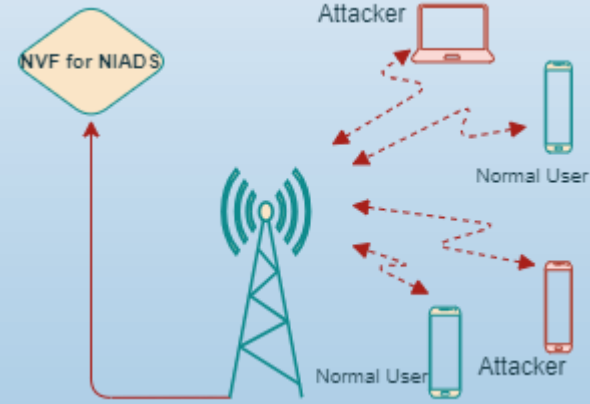
- Private Campus Networks (PCNs) are critical in providing customized connectivity for industries such as smart factories and autonomous driving.
- Technology growth brings increased security risks along with flexibility.
- The increasing complexity of network attacks has made it necessary to develop more sophisticated intrusion detection systems to detect and respond

Related Works

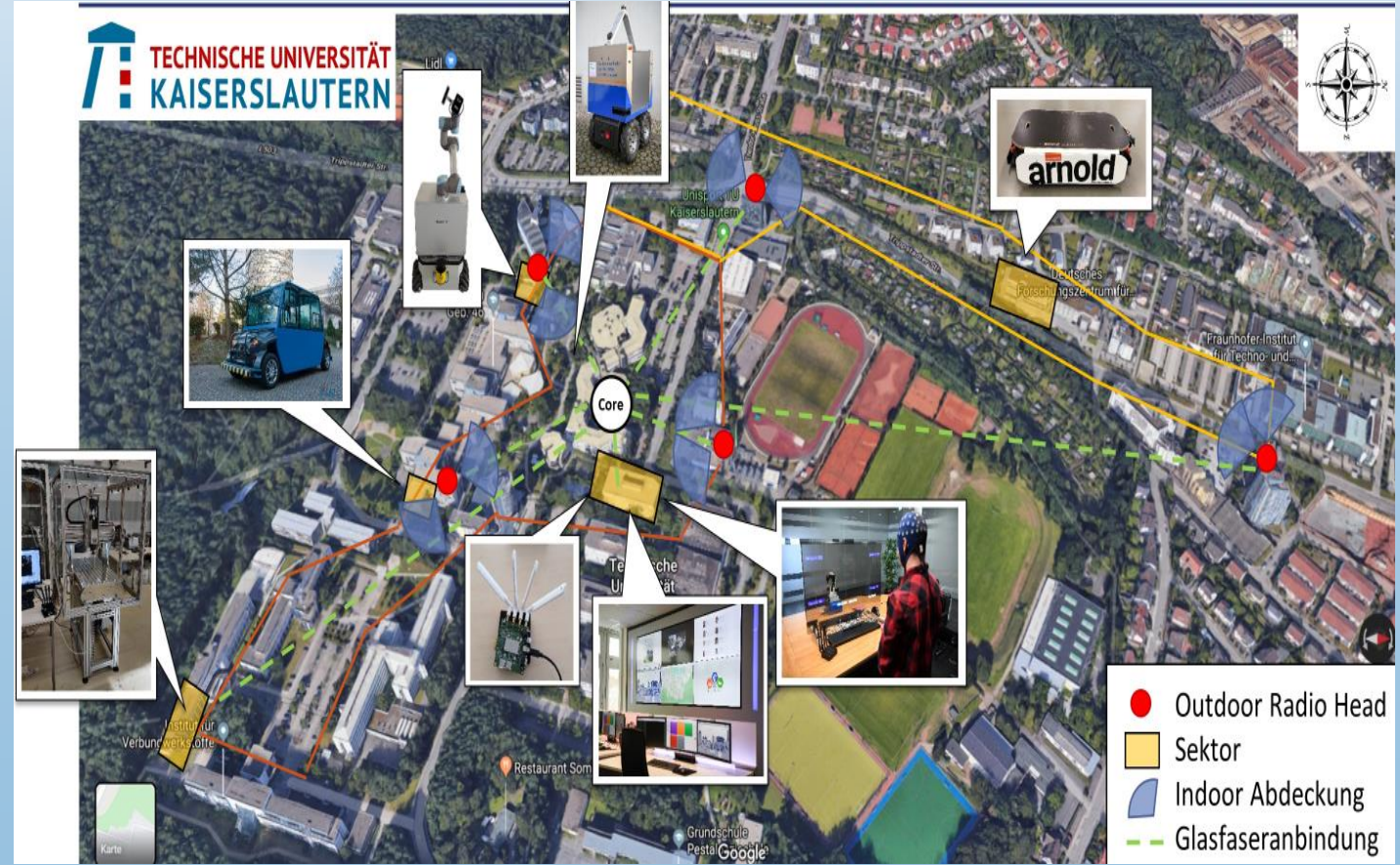
Authors	Contribution	ML model	Database
Almiani et.al - 2021	impact of 5G on IoT networks and proposed a DDoS intrusion detection model which utilize a deep Kalman back propagation neural network	Kalman back propagation neural network.	CICDDoS2019
Alimi et.al - 2022	Intrusion Detection systems for IoT	RLSTM	CICIDS-2017 and NSL-KDS
Wang et.al - 2023	multi-class network traffic classification but the authors concluded that there are minimal accuracy improvements at the cost of very high inference time for the combinations	DNN, CNN, RNN, LSTM, and their combinations	CSE-CIC-IDS2018
Sood et.al - 2023	a two-stage network traffic anomaly detection	not mentioned	UNSW-NB15

- **Our contribution**

- To enhance the robustness of IDS, we have generated a dataset from multiple private campus 5G networks. NIADS - Autoencoders and LSTM Autoencoders



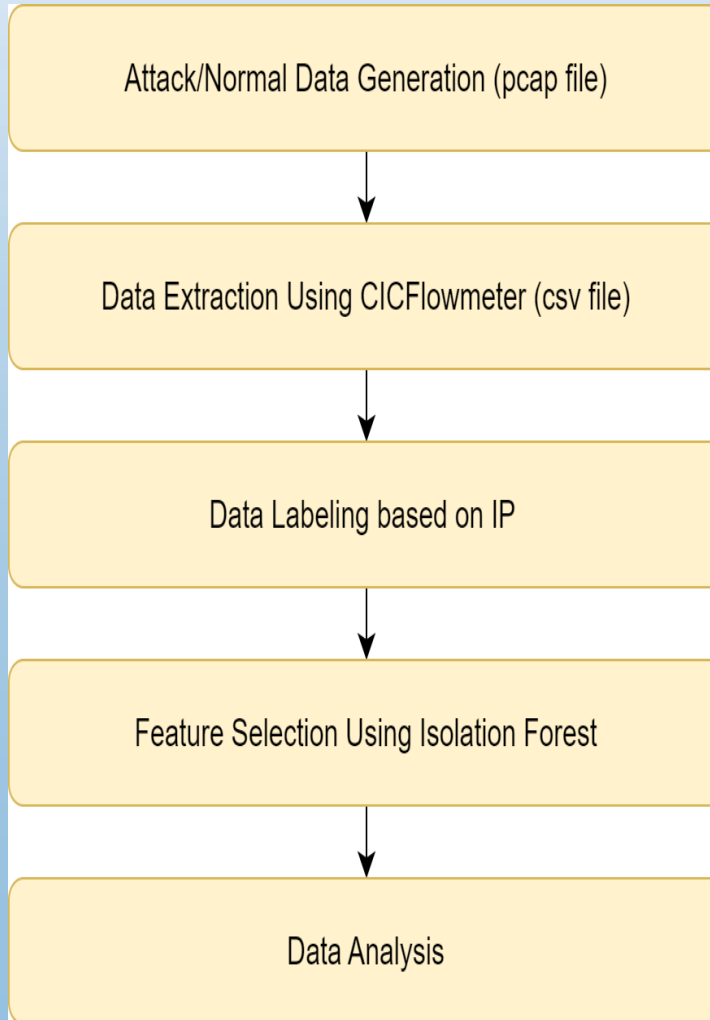
Network parameters	Value
Frequency range	3.7-3.8 GHz (100 MHz Bandwidth)
Indoor PCN	Nokia, Mecsware, Amarisoft
Outdoor PCN	Nokia
Normal UE's	Quectel RM500Q-GL, Samsung S23
Attacker UE's	Rpi 4 with Kali linux + Quectel RM500Q-GL



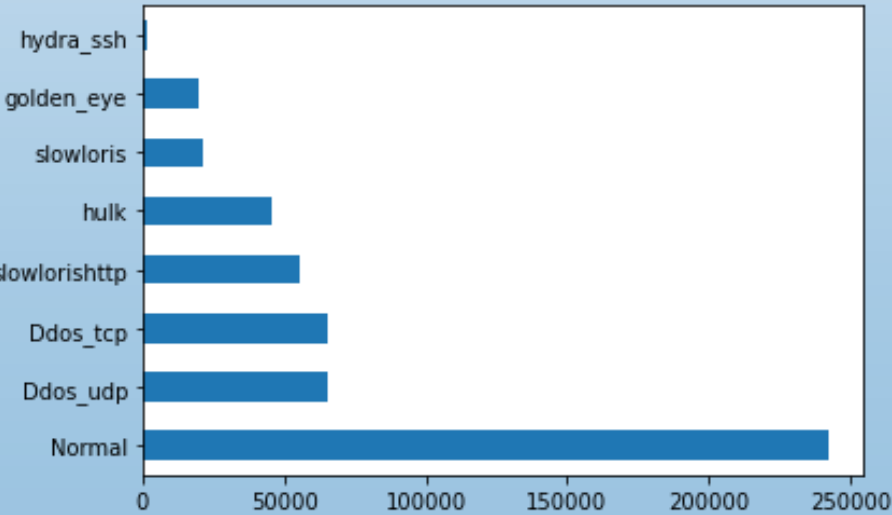
Legend:

- Outdoor Radio Head
- Sektor
- Indoor Abdeckung
- Glasfaseranbindung

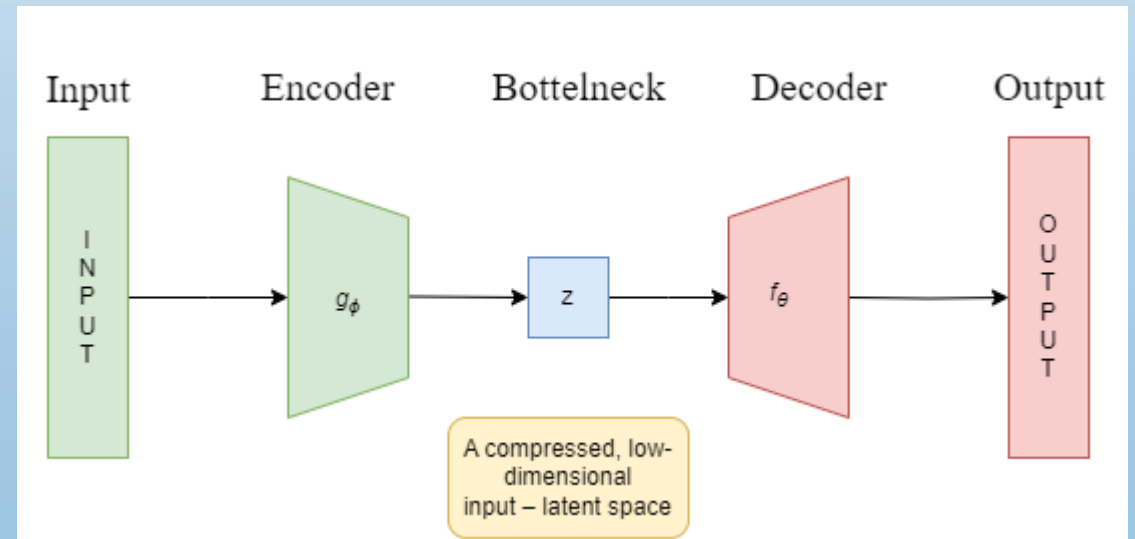
DATA COLLECTION PROCESS



- 5,33,521 tuples
- Normal to Malicious data - 4:1
- Isolation forest method is used to isolate outliers – 34,622
- 33 independent features



- 2 deep learning models were used
 - Autoencoders and LSTM Autoencoders
- 7:3 data split for training and testing
- Confusion Matrix, AUC, Precision, Recall, F1-Score, Macro Average, and Weighted Average are used to indicate performance



Autoencoders

- Trained with 100 epochs
- AUC 86 %

Metric	Precision	Recall	F1-Score	Support
0.0 (Normal)	0.83	0.87	0.85	63301
1.0 (Abnormal)	0.90	0.86	0.88	81269
Accuracy			0.86	144570
Macro Average	0.86	0.87	0.86	144570
Weighted Average	0.87	0.86	0.86	144570

Label	Normal	Abnormal
Normal	85.9 %	14.1 %
Abnormal	12.8 %	87.2 %

LSTM Autoencoders

- Trained with 50 epochs
- AUC 96 %

Metric	Precision	Recall	F1-Score	Support
0.0 (Normal)	0.63	0.92	0.75	14642
1.0 (Abnormal)	0.99	0.96	0.97	175618
Accuracy			0.95	190260
Macro Average	0.81	0.94	0.86	190260
Weighted Average	0.97	0.95	0.96	190260

Label	Normal	Abnormal
Normal	91.7 %	8.3 %
Abnormal	4.4 %	95.6 %

Autoencoders

- Trained with 100 epochs
- AUC 86 %
- The autoencoder's performance is good, but there is room for improvement
- The recall for the abnormal class could be improved by adding more malicious training datasets or by adjusting the autoencoder's hyperparameters.
- Standard autoencoders might not be able to capture long temporal dependencies

LSTM Autoencoders

- Trained with 50 epochs
- AUC 96 %
- LSTMs allow the long temporal dependencies to capture the order and flow of network data

- Dataset generated from three different PCN 5G SA
- Two ML-based NIADs for PCN
 - Autoencoders and LSTM autoencoders
- LSTMs allow the long temporal dependencies to capture the order and flow of network data
- LSTM Autoencoders perform better with an accuracy of 96%, whereas autoencoders with 86% accuracy
- Limitations :
 - Models are still vulnerable to zero-day attacks ,need a large range of data to understand the pattern effectively
- Future work:
 - Focus on Generative Adversarial Networks (GAN) models as they can create synthetic data that resembles training data.
 - GAN models highlight the potential of data augmentation

Thank you