Abstract 2 (2022-10-10)
Anomaly Detection in IOT Devices using LogBERT
Speaker: Eniela Vela

Speaker discussed about the IOT devices usage across globe. According to her analysis, around 20 % of the companies using IOT devices have reported an cyberattack in 2020. Also, in 2022, it reaches to approximately 10 billion and projected that it will reach around 30 billion in 2030. She describes how machine learning helps to provide security in IOT devices. For example, Firstly, speaker discussed that K-nearest neighbor (KNN) has achieved 84.82% compared to decision tree (81.05%). Secondly, SVM performed 100% accuracy using blackhole and sinkhole attacks. Further she discussed about the unsupervised machine algorithm LogBert which has very good result on three different datasets like Hadoop, BlueGene and thunderbird. She described three question that relevant to her research like can logBERT usage can be extended in IOT device anomaly detection, the performance of the algorithm and measures of indication on anomaly detection.

Further speaker discussed about the anomaly detection with the example where she took an example of dataset and explained value student is different from values that are present in the dataset. LogBERT is equal to LogParser + BERT where LogParser contains drain (extract log templates and automatically split them into the group) and spell (find matches across different logs) and BERT contains masked language modeling (MLM - that predicts original value from the context) and next sentence prediction (NSP – recognizes the second sentences in a pair of sentences in the following sentences in the original documents). LogBert using masked log key predication and volume of hypersphere minimization.

Further speakers describe about the data generation where she took data from different routers. The routers have indoor and outdoor camera, smart lightbulb, and DVR. All the attacks were performed in the virtual box and Arduino to check the electric current, Wireshark to get the network traffic data. She discussed about the usage of three malware for data generation. Moreover, she discussed about the logBERT how it will understand the patterns like normal or abnormal. She generated over 9 million data set. Further, she described about the data analysis and for this experiment she used network traffic data. For data cleaning, she removed all the missing data from the dataset. She did the grouping based on the same source and destination IP and protocol. For the visualization, speaker created the UML diagram to better understand the work or process flow.

Further speaker described about the mini set to the test the balances of data which had result into 70% normal and 30% abnormal. The algorithm was trained in normal dataset and F1- score is 55.56%. Also, train loss and valid loss described in the line graph. Another mini dataset speaker took 90% normal and 10% abnormal and F1- score was 75.5%. Afterwards she took the comparison of computer devices and logBERT where indoor camera had the best results. For three IOT devices logBERT had better result and other three had logCluster had better result. Further, speaker calculated the energy consumption and found that when attacks happened energy consumption is twice as normal state.

For future work, speaker suggested she will be improving logBERT algorithm data features, incorporate evaluation for machine leaning and evaluate performance of logBERT in high data.