**THESIS FINDINGS**

**Proposed Architecture-**

* **Synthetic Datasets Collection**:
  + The testbed architecture consists of multiple host machines connected like the Internet. These machines generate realistic background traffic based on benign and malicious simulations.
  + The extracted data closely resembles real-world data in Packet Capture (PCAP)s.
  + Data is engineered using network analysis tools for feature extractions and saved as raw CSV files. These representations are used to study intrusions in a network system.
* **NSL-KDD Dataset**:
  + NSL-KDD is a widely studied dataset for evaluating machine learning-based Intrusion Detection Systems (IDSs).
  + Label encoding is opted, iterating on all categorical features during the preprocessing pipeline and save the encoded values as a JSON object in the utils directory. This utility file is later used during model training and also while crafting adversarial examples.
  + It has limitations like incomplete traffic, duplicate records, and outdated attack types.
  + New DoS/DDoS datasets have been proposed by CIC to address these limitations.
* **DDoS Attacks**:
  + DDoS attacks are a significant threat to network security, targeting layers [3-7] of the OSI model.
  + An attacker uses multiple malware-infected devices in a distributed fashion, known as a botnet, controlled by master nodes.

**Baseline models:**

* **Five machine learning-based models** are used for experiments:
  + Random Forest
  + Linear SVC
  + Bernoulli Naive Bayes
  + Gradient Boost
  + Neural Network
* Random Forest and Neural Networks outperform other models like SVC, Naive Bayes, and Gradient Boost.

**Adversial Attack Algorithm**

**C-LowProFool**

* **Objective**: Focuses on imperceptibility and functionality while evading a trained model.
* **Mechanism**: Uses a combination of cross-entropy loss and perceptibility loss, controlled by a hyper-parameter *λ*
* **Performance Metrics**:
  + NSL-KDD: Success Rate 0.906, Attack Score 80.7, Perturbation Distance 0.327
  + CICDDoS: Success Rate 0.437, Attack Score 68.9, Perturbation Distance 0.336
  + CICIDS: Success Rate 0.729, Attack Score 45.2, Perturbation Distance 0.13

**C-DeepFool**

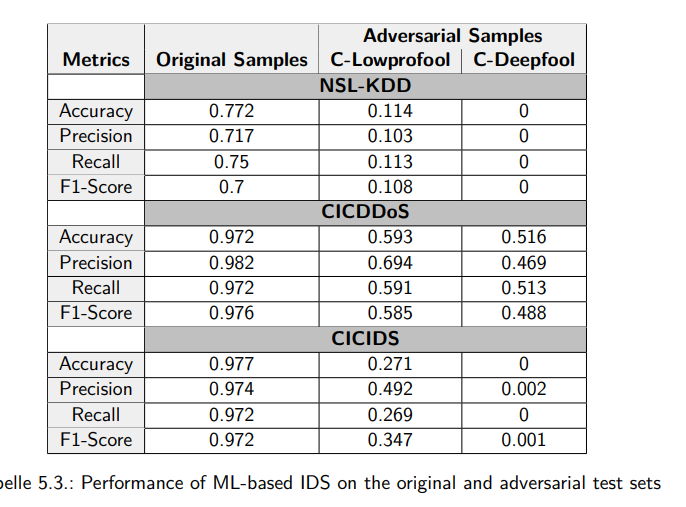
* **Objective**: Aims to minimize the L2​ norm of the perturbation *δ* without any constraint on perceptibility
* **Mechanism**: Uses iterative linearization of the victim model to generate perturbations.
* **Performance Metrics**:
  + NSL-KDD: Success Rate 1, Attack Score 94.3, Perturbation Distance 0.576
  + CICDDoS: Success Rate 0.493, Attack Score 199.4, Perturbation Distance 0.755
  + CICIDS: Success Rate 0.1, Attack Score 91.9, Perturbation Distance 1.26

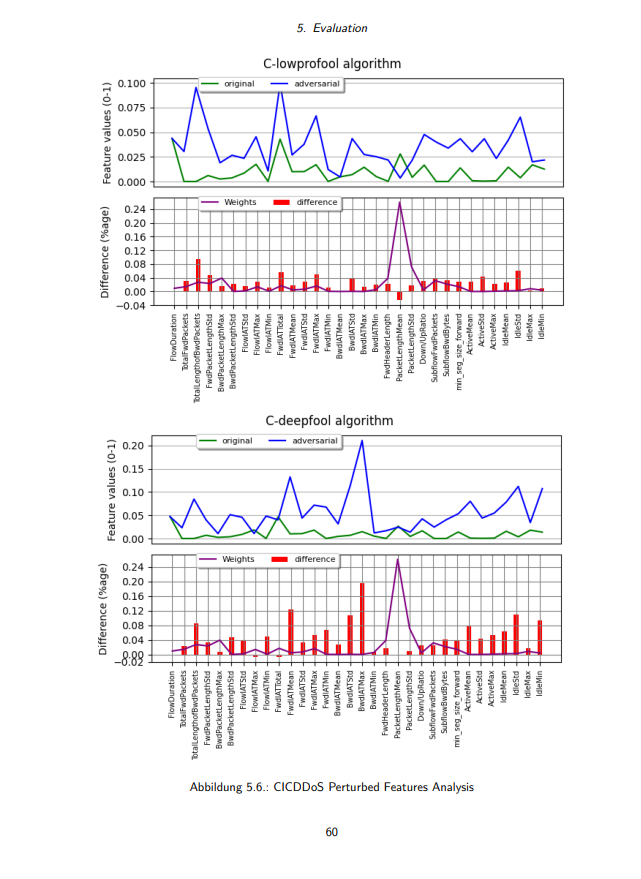
**Comparison**

* C-DeepFool generally has a higher success rate and attack score but at the cost of less realistic perturbations compared to C-LowProFool

**Evaluation:**

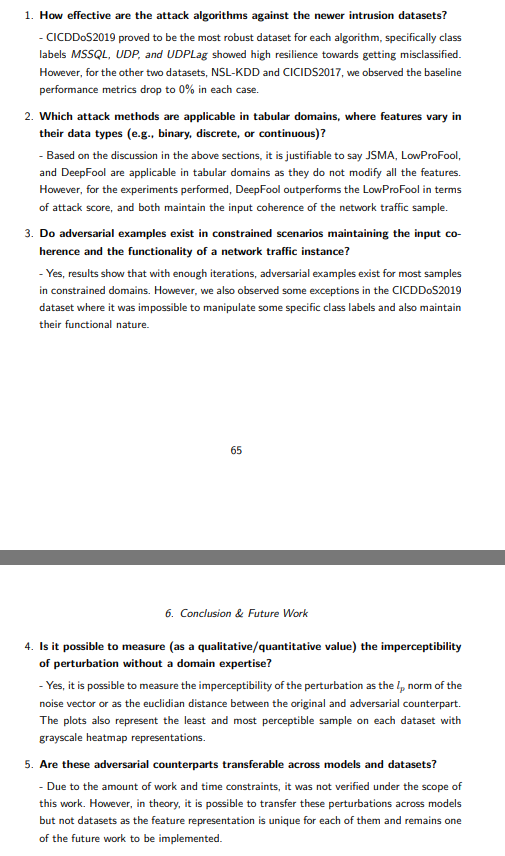
* **CICDDoS Perturbed Features Analysis**:
  + Below figure Uses C-lowprofool, c-deepfool, mentioned that pertains to CICDDoS Perturbed Features Analysis.





**Conclusion & Future Work**

* **Imperceptibility of Perturbation**:
  + It's possible to measure the imperceptibility of the perturbation as the lp norm of the noise vector or as the Euclidean distance between the original and adversarial counterpart.
* **Transferability of Adversarial Counterparts**:
  + In theory, it's possible to transfer these perturbations across models but not datasets due to unique feature representation.



* **Future Work**:
  + Investigate why some class labels in CICDDoS2019 showed high robustness towards attack algorithms.
  + Explore the concept of transferability as an extension to this research work.
  + Find relevant defense algorithms to prevent attacks and build a reliable networked system.