Heterogeneous Network Inspection in IoT Environment with FPGA based Pre-Filter and CPU based LightGBM

Zhenguo Hu[1], Hirokazu Hasegawa[2], Yukiko Yamaguchi[1], Hajime Shimada[1]

Nagoya University[1]

National Institute of Informatics[2]

2023/9/28

名古屋大学 NAGOYA UNIVERSITY

Contents

- Background
- Heterogeneous Malicious Traffic Detection System Design
  - FPGA based Pre-Filter
  - Machine Learning based Traffic Detection
- Experiment & Evaluation
  - Evaluation on Training Stage
  - Evaluation on Inference Stage
- Conclusion

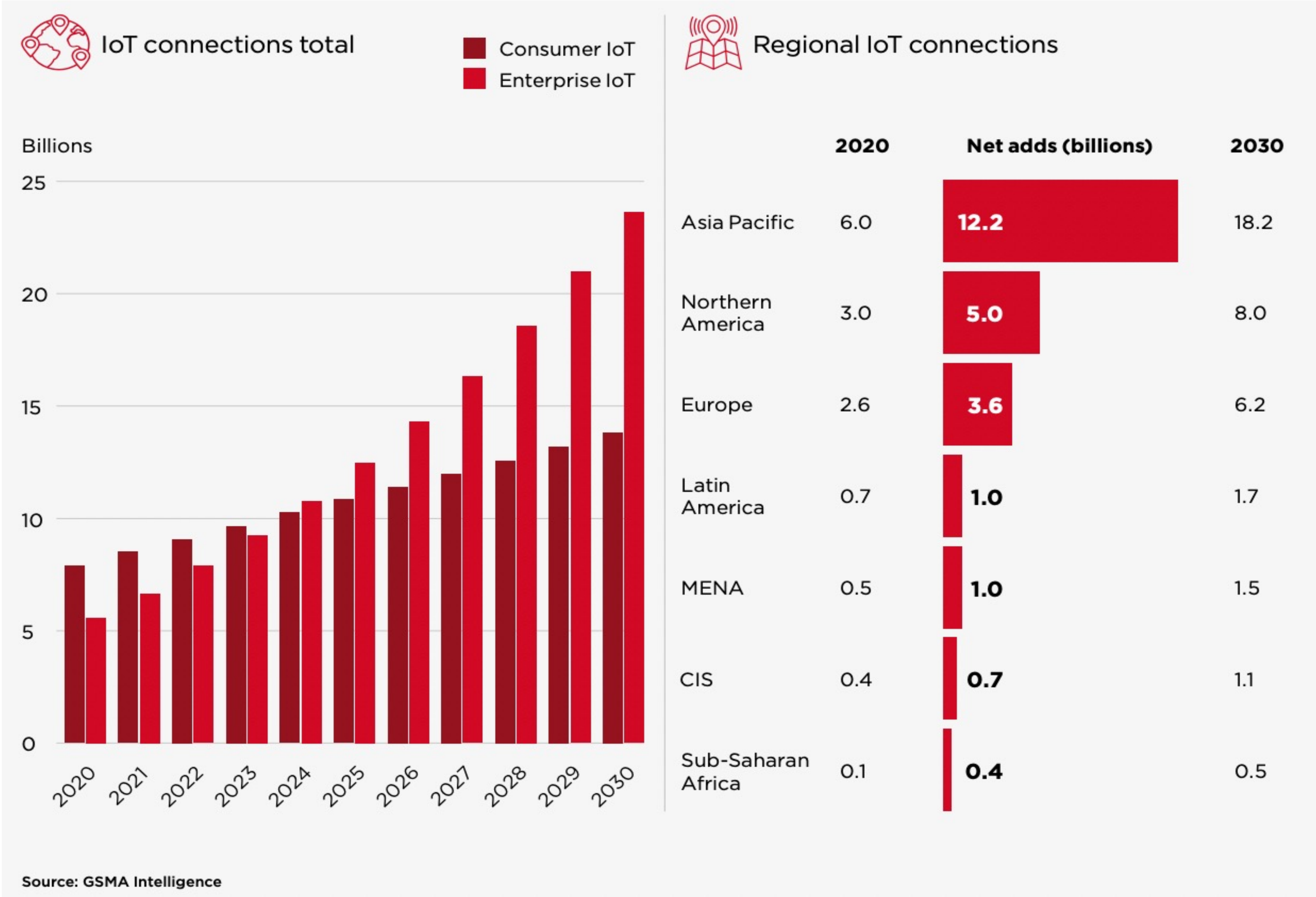
Contents

IoT Development

- Nowadays, with the development of modern society, IoT has entered many aspects of our daily lives.

  - Both consumer IoT connections and enterprise IoT connections will increase in recent years.

  - IoT connections are predicted to experience the majority growth in the Asia Pacific region.



[1]

[1] IoT for Development: Use cases delivering impact, GSMA Intelligence

名古屋大学 NAGOYA UNIVERSITY

Cyber Attack

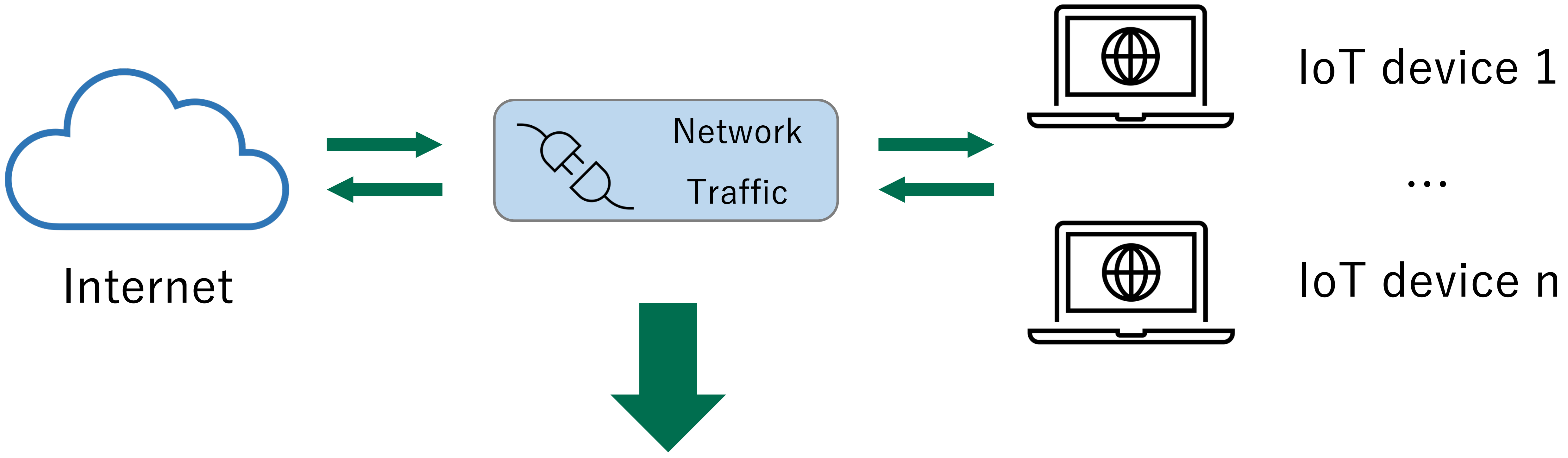
- With the development of IoT technologies, IoT devices and connections will suffer many various malicious attacks on privacy and security.

  - Bruteforce Attack
  - DDoS Attack
  - SQL Injection Attack
  - …

How to protect the network security and detect malicious traffic under IoT environment has become the common goal of researchers in the whole world.

NIDS

- NIDS is a hardware device or software application which is deployed to identify network threats.
  - Malicious attacks against the IoT network are gradually increasing.
  - NIDS can be deployed to execute complex malicious traffic detection.

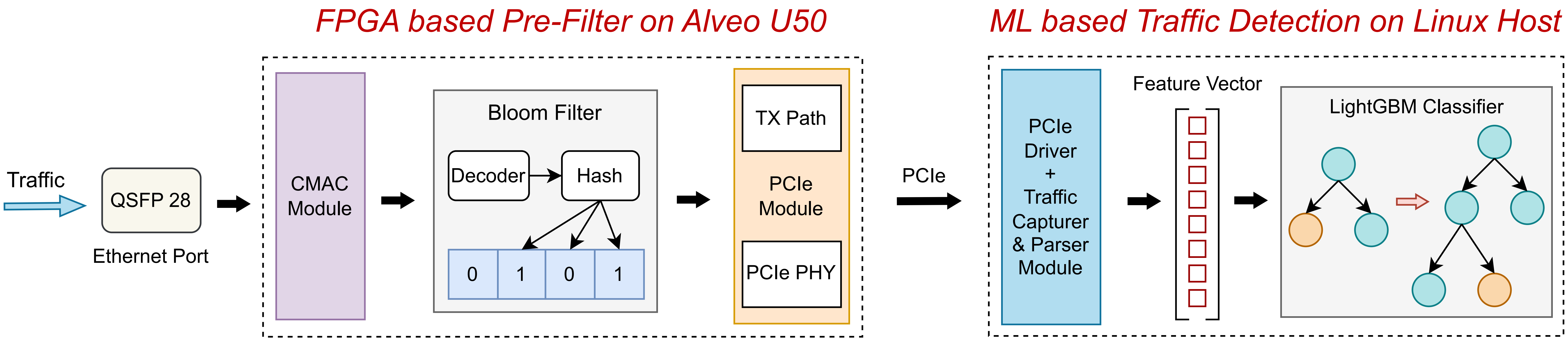


In order to detect malicious traffic especially in IoT environment, it is important to build an effective defense system to prevent attacks.

Contents

Heterogeneous Malicious Traffic Detection System Design

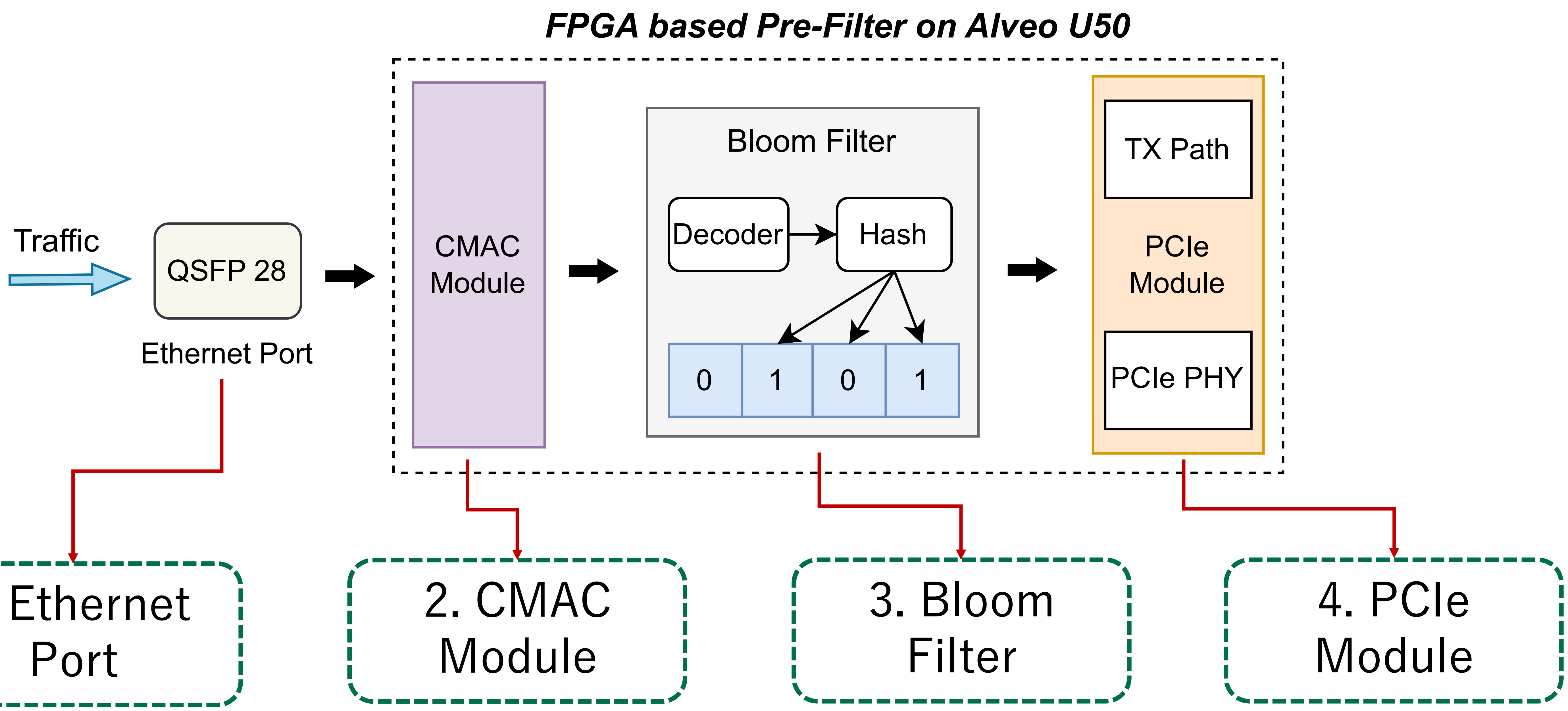
- The heterogeneous malicious traffic detection system is designed to detect packet-level malicious traffic.



- It mainly consists of two parts:
  - FPGA based Pre-Filter
  - Machine Learning based Traffic Detection

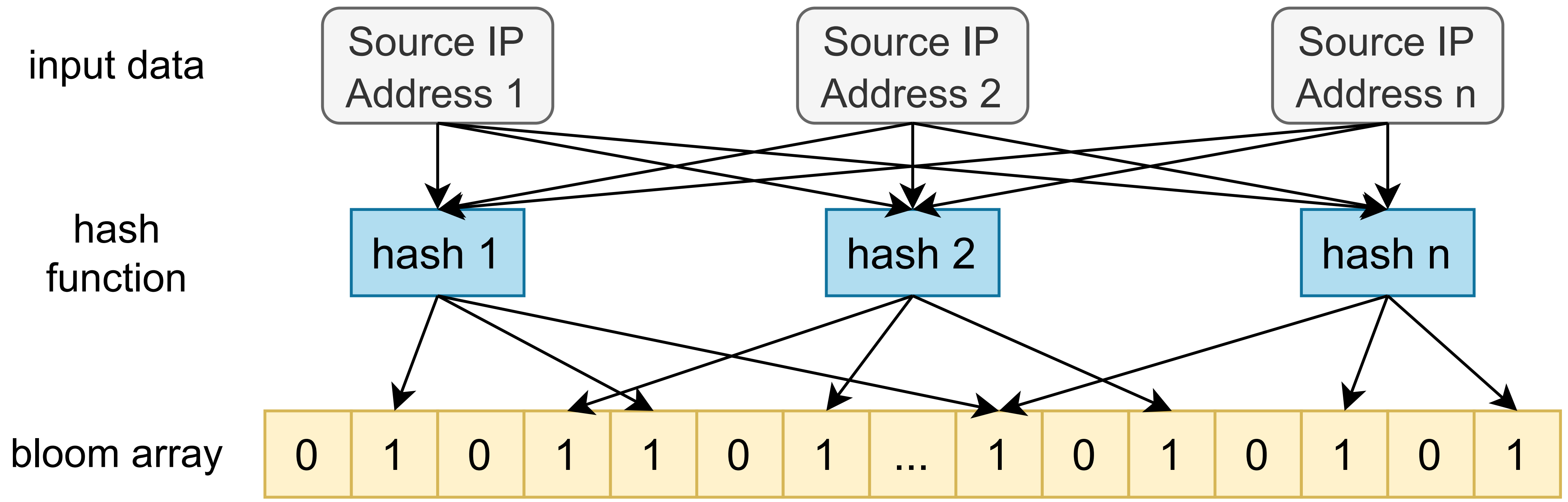
FPGA based Pre-Filter

- The FPGA based Pre-Filter is used to filter the truly malicious traffic by setting a blacklist in the FPGA board.



FPGA based Pre-Filter (Bloom Filter)

- Bloom filter is used as the IP blacklist implementation and to filter the “Source IPv4 Address”.
  - Cyclic Redundancy Check (CRC) hash function
  - $False\ Positive\ Probability = (1 - e^{\frac{-kn}{m}})^{k}$

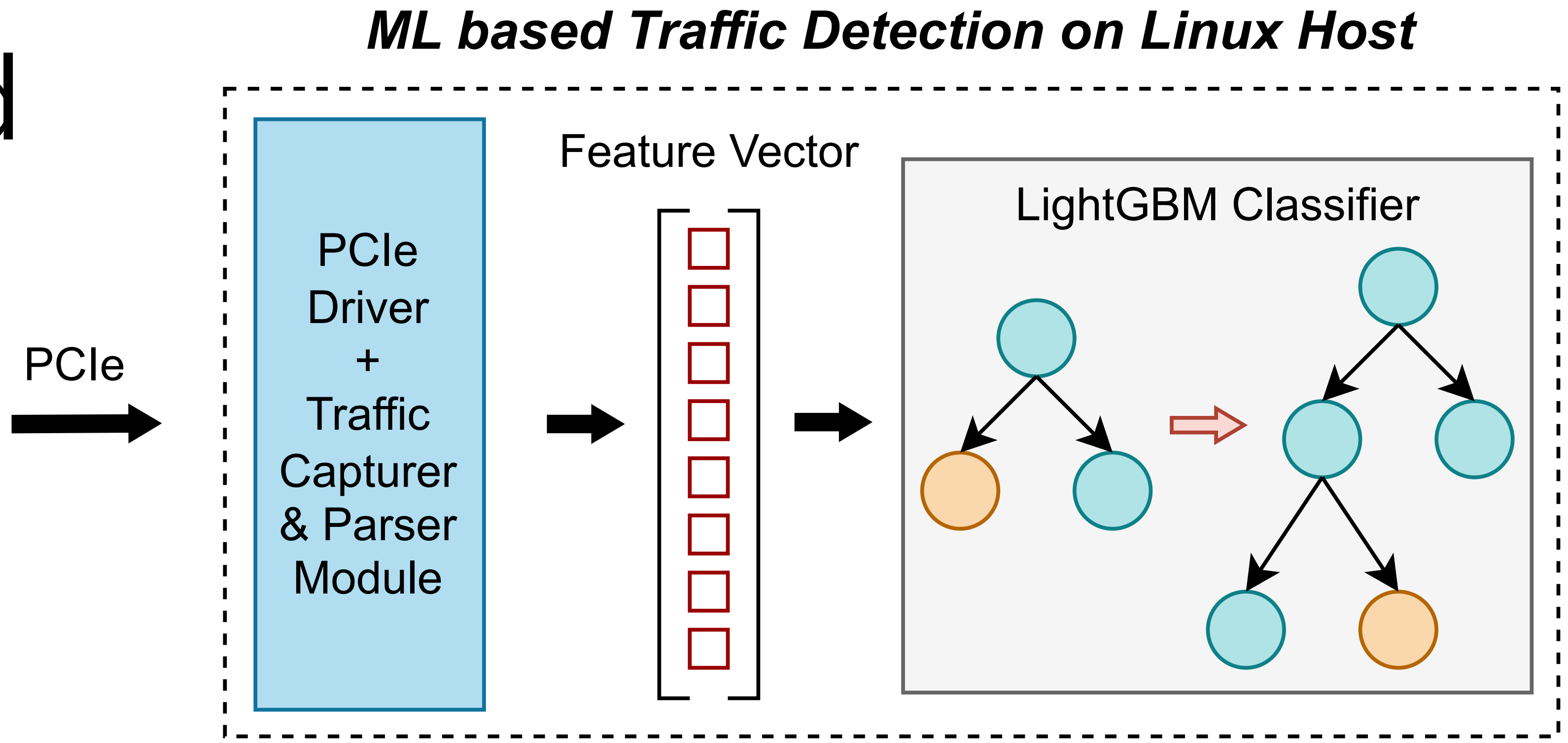


- K=4, m/n=16, false positive probability = 0.239%

Machine Learning based Traffic Detection

- Machine learning based traffic detection leverages the data-driven insight ability of machine learning to analyze the malicious traffic on the CPU side.

- Two modules are implemented to detect the attack behaviors:
  - Traffic Capturer and Parser Module
  - LightGBM Classifier



Machine Learning based Traffic Detection

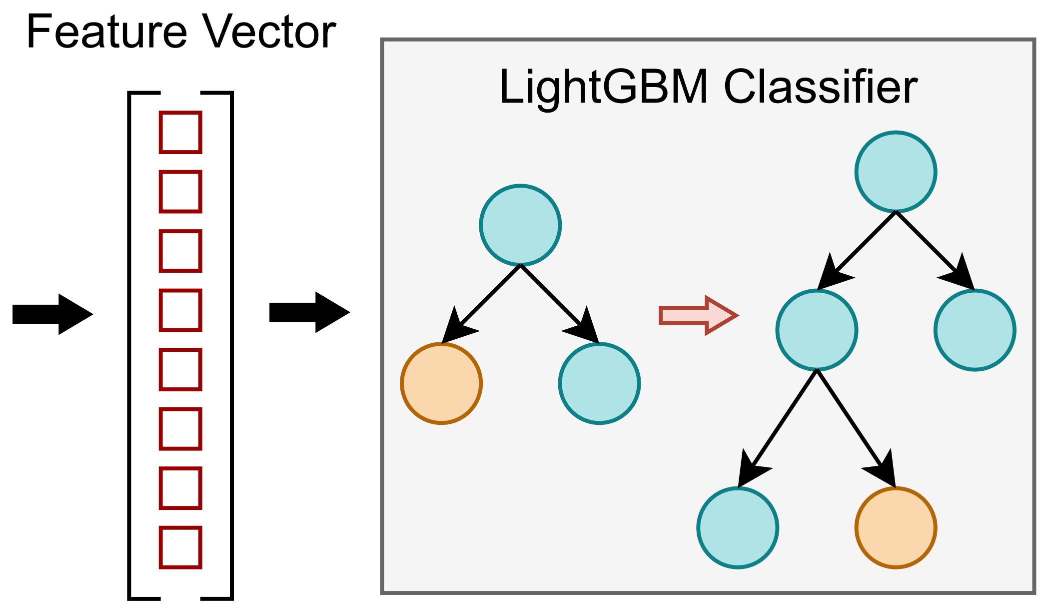
Feature Extraction

- We focus on extracting packet-level traffic features.
- We encode the label for Layer, Source IP Address and Destination IP Address to convert the category format into the number format.

Feature	Description
IPv4 Length	The length of an IPv4 packet
IPv4 ID	The identification of an IPv4 packet
IPv4 TTL	The time to live of an IPv4 packet
Layer	The type of protocol
Source Port	The source port
Destination Port	The destination port
Source IP Address	The source IP address
Destination IP Address	The destination IP address

LightGBM Classifier

- We train a LightGBM model in advance and and then instantiate it as the classifier implementation.



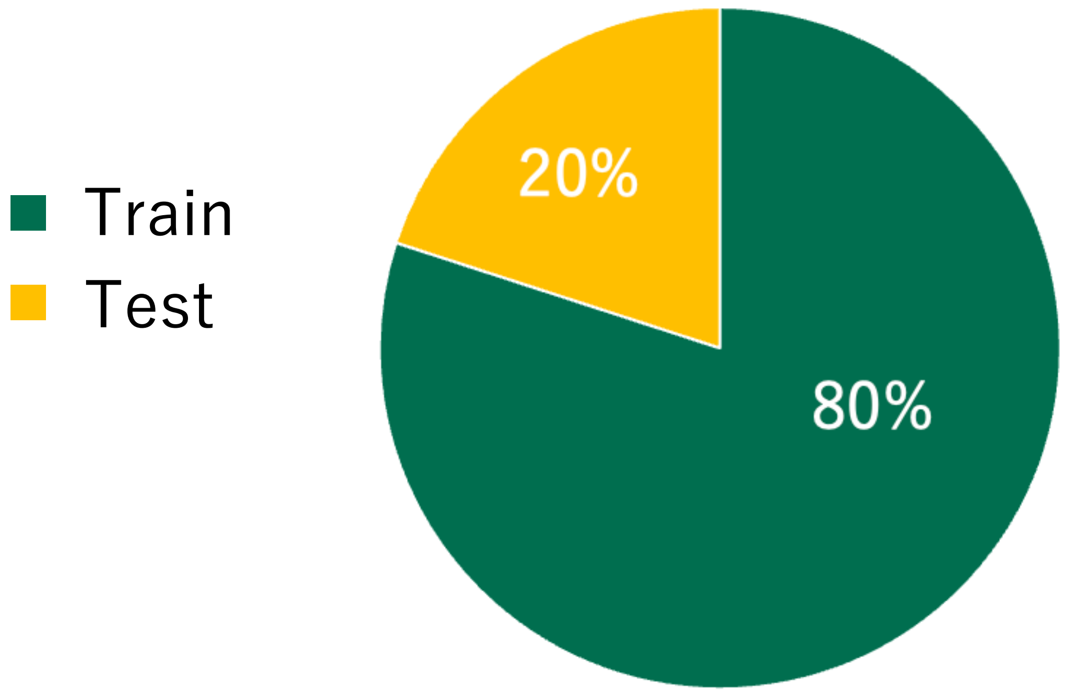
- The extracted features are combined into a feature vector and sent to the LightGBM to execute the prediction.

Contents

Experiment & Evaluation (Training Stage)

Dataset

- The experiment and evaluation dataset comes from Kitsune Mirai.
- It is presented in 2018 and it is captured from an IoT network, where the Mirai malware begins to infect other devices and scans for new victims network.



Number	
Num of Malicious	*Num of Benign*
642,516	121,621
Num of Train	*Num of Test*
611,309	152,828

- It consists of 642,516 pieces of malicious data and 121,621 pieces of benign data. We select 80% (611,309) as the training set and 20% (152,828) as the test set.

Experiment & Evaluation (Training Stage)

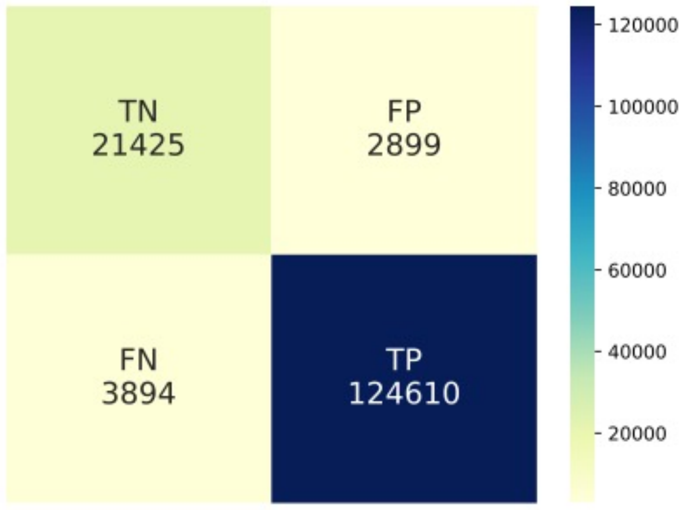
Evaluation Metric

- False Positive (FP), False Negative (FN), Accuracy (ACC), Precision, Recall, F1-score

Classifier	ACC	Precision	Recall	F1-score
SVM	0.9104	0.9706	0.9214	0.9453
KNN	0.9494	0.9740	0.9656	0.9698
Random Forest	0.9525	0.9766	0.9668	0.9716
Decision Tree	0.9556	0.9773	0.9697	0.9735
LightGBM	**0.9589**	**0.9774**	**0.9736**	**0.9755**

Evaluation Results

- We set four other models including Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest and Decision Tree as the comparisons.
- LightGBM achieves the highest evaluation in all the classifiers on ACC, Precision, Recall and F1-score.



(a) Decision Tree

(b) LightGBM

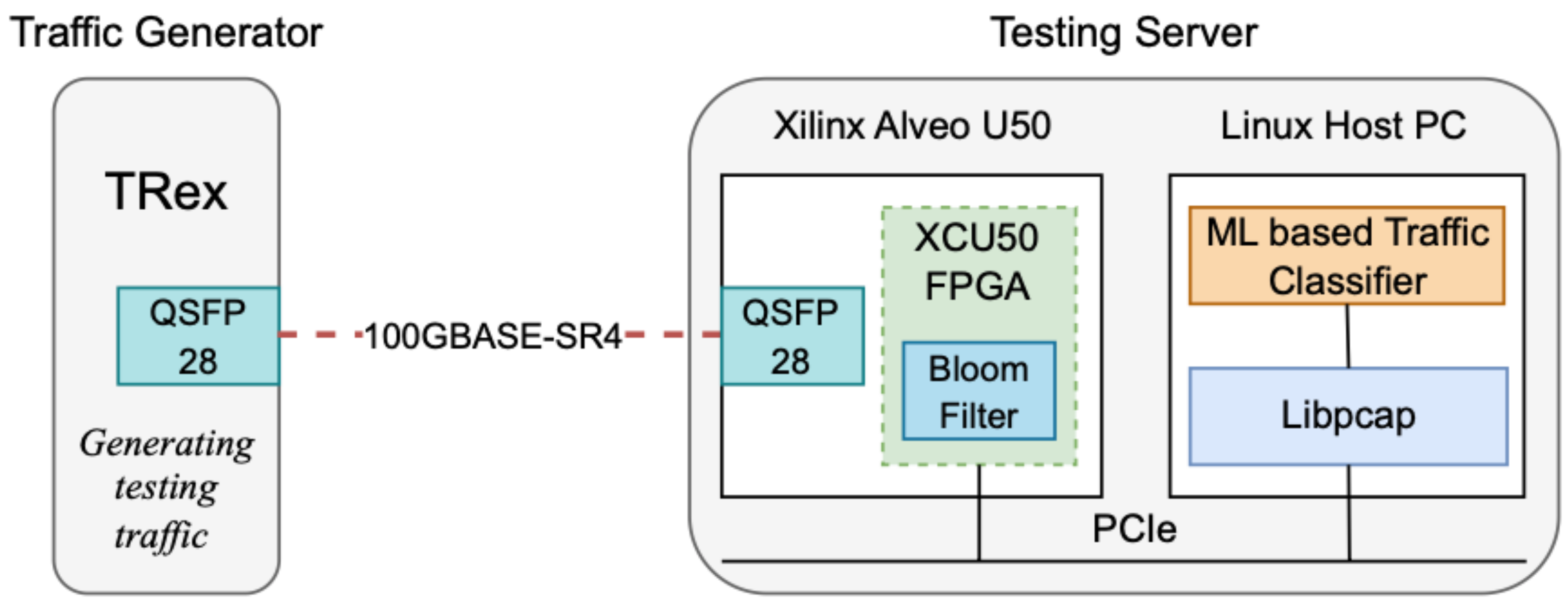
Experiment & Evaluation (Inference Stage)

Experiment Environment

- The assumed experimental IoT network
  - By adjusting the bloom array in the FPGA based pre-filter, we build four situations of irregular traffic which includes different types of source IPv4 addresses.

Source IPv4 Address Range	**Clients**	**Assumed Irregular Traffic filtered by FPGA based Pre-Filter**	
		Source IPv4 Address	***Situations***
192.168.2.0/24 and 0.0.0.0	30	192.168.2.108	Situation1
		192.168.2.108 192.168.2.1	Situation2
		192.168.2.108 192.168.2.1 192.168.2.113	Situation3
		192.168.2.108 192.168.2.1 192.168.2.113 192.168.2.110	Situation4

- The experiment environment setup
  - Cisco TRex is used as the traffic generator to generate and send the traffic to the testing server.
  - On the testing server, the Xilinx Alveo U50 accelerator card is our FPGA platform.



Experiment & Evaluation (Inference Stage)

Evaluation Results

FPGA Resource Utilization

- The resource consumption of LUT, Register and BRAM Tile comes from the Xilinx Vivado Design Suite 2020.2.

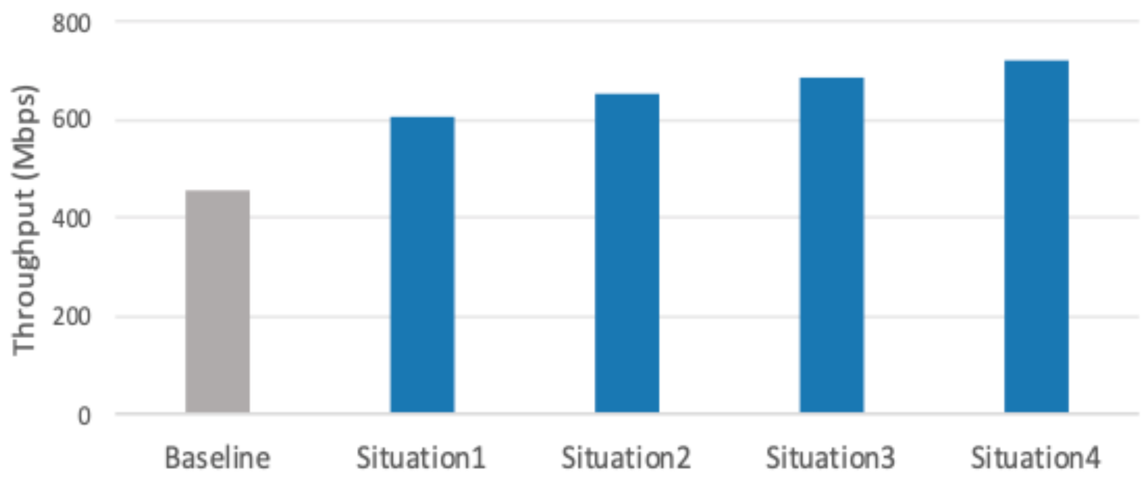
- From the result we can see, the filter module consumes 5,279 LUTs and 5,209 registers which occupies a reasonable resource consumption.

Module Name	LUT	Register	BRAM Tile
CMAC Module	9,793	31,504	0
Filter Module	5,279	5,209	0
PCIe Module	79,576	84,544	94
Proportion	10.9%	7.0%	7.0%

Throughput

- We evaluate the throughput under various situations. Baseline indicates that there are no rules being enabled in the pre-filter.

- The FPGA based pre-filter can block a portion of malicious traffic which reduces the burden on ML based traffic detection, and improve the detection performance.



Contents

Conclusion

- We present a realtime heterogeneous malicious traffic detection method using FPGA based pre-filter and machine learning based traffic detection especially in IoT environment.

- We design an experiment to evaluate the proposed system and the results show that it has better performance with low FPGA resource usage and effective throughput improvement.

- In the future, we will explore to add other models to the system to further improve the detection efficiency.

Thank you for your listening.