Image-based Neural Network Models for Malware Traffic Classification using PCAP to Picture Conversion

Georgios Agrafiotis[1], Eftychia Makri[1], Ioannis Flionis[1], Antonios Lalas[1], Konstantinos Votis[1], Dimitrios Tzovaras[1] and Nikolaos Tsampieris[2]

[1]Information Technologies Institute / Centre for Research and Technology Hellas
6th km Xarilaou – Thermi, Thessaloniki, 57001, Greece

[2]Infitheon Technologies
Patriarchou Grigoriou & Neapoleos 27, Athens, 15310, Greece

5-8 September 2022 ICICS Canterbury, UK 2022

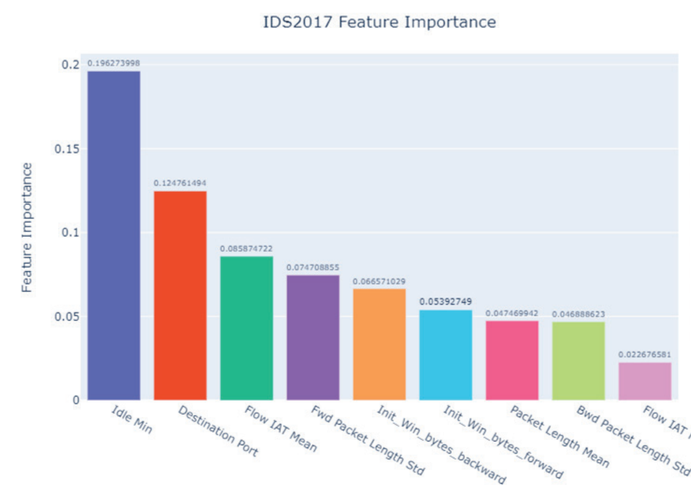
Background

- The 5th generation (5G) networks are deemed a rapidly emerging technology with multiple applications and associated **cybersecurity challenges**, rendering **traffic categorization** a task of paramount importance in the network security sector, as well as the first stage in a network-based **intrusion detection system** (IDS).
- However, current detection systems, which work better when malware signatures are known in advance, are useless against **zero-day assaults** and **amorphous malware** and unable to deal with developing malware vectors or emerging network protocols, e.g. the **5G-related protocols**.

Goal

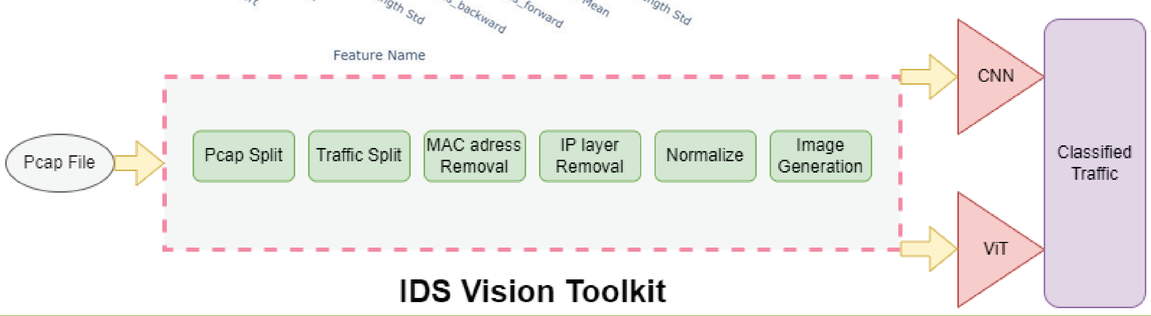
- Transformation of **raw traffic data** into grayscale **pictures** and employment of state-of-the-art image-based neural network models, namely **Vision Transformers** (ViT) and **Convolutional Neural Networks** (CNN) for an **IDS application properly tailored on 5G**.

Data Preprocessing



- CIC-Flow-Meter for the extraction of **80 statistical network traffic features**
- Clearance of redundant records
- Discard of extra features

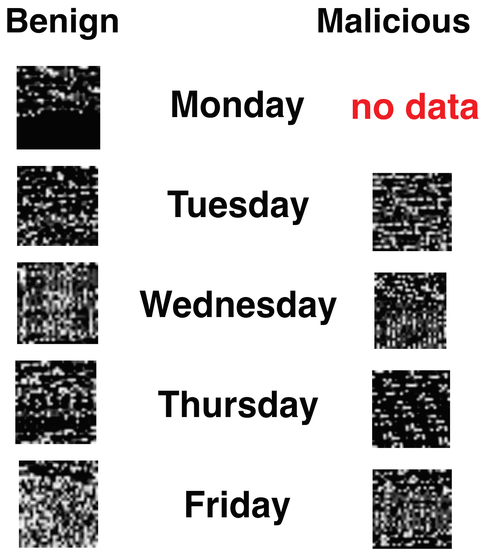
Comparison of **CIC-Flow-Meter** and **IDS Vision Toolkit** means informative comparison between **statistical** and **temporal data**



With **IDS Vision Toolkit**, raw traffic data (PCAP) is converted either png or jpg format

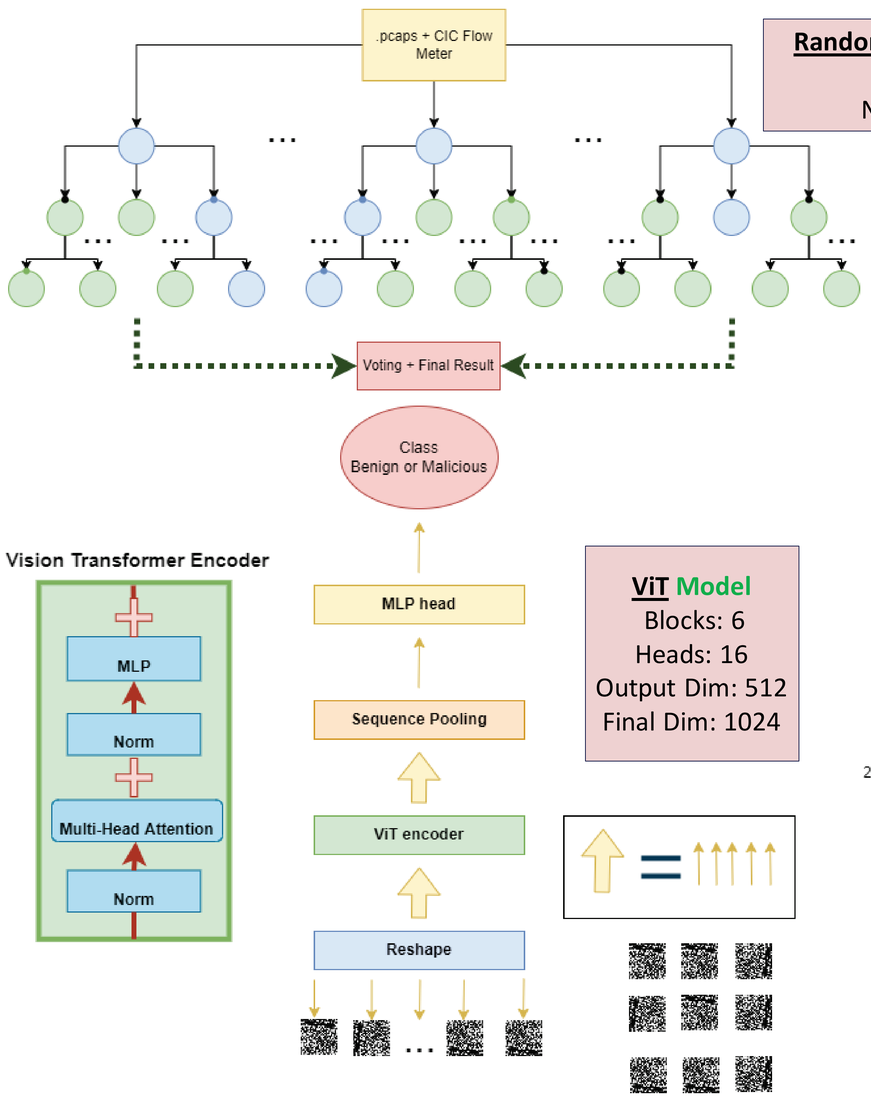
- **Step 1: Traffic Splitting**
- **Step 2: Traffic Clearing**
- **Step 3: Image Generation**

Data Visualization



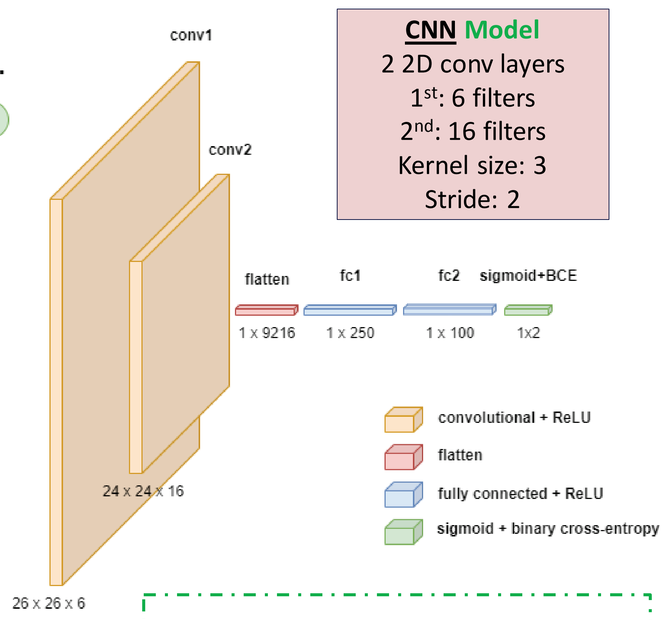
- **4,834,730** images, 4,046,667 benign and 788,053 malicious
- Each image represents a **flow**
- Each greyscale image is **28 × 28** bytes

Models' Architecture



Random Forest (RF) Model
Depth: 7
No. of trees: 78

ViT Model
Blocks: 6
Heads: 16
Output Dim: 512
Final Dim: 1024



CNN Model
2 2D conv layers
1st: 6 filters
2nd: 16 filters
Kernel size: 3
Stride: 2

To the best of our knowledge, **ViT** models **have not been examined** yet in terms of **cybersecurity** applications.

Experimental Results

CIC-Flow	RF	IDS Vision Toolkit	CNN	ViT
Accuracy	0,985	**Accuracy**	0,99	0,999
Precision	0,979	**Precision**	1	1
Recall	0,931	**Recall**	0,997	0,998
F1	0,954	**F1**	0,998	0,999

- **IDS-Vision** toolkit pre-processing method, leads to noticeable **better results** in comparison to the **CIC-Flow-Meter**
- **ViT** and **CNN**, achieved **almost the same** performance on the image dataset.

Conclusions

- The proposed preprocessing toolkit, named **IDS-Vision Toolkit**, extracts **images from PCAP files**, for classification of benign and malicious network traffic and it is a **promising tool for the employment of representation learning** as a technique for IDS purposes.
- The two experiments, in terms of different datasets (**statistical** and **tabular**) and ML algorithms exhibited promising results for the aforementioned classification task, achieving **accuracy higher than 98%**.
- The **image dataset** (tabular) as an input, led to **better results**, in comparison with the **tabular** dataset, paving the way for the opportunity to utilize **state-of-the-art models for image recognition tasks** for IDS purposes.
- **Future work** includes testing against **a larger and 5G-specific dataset** and on a **multi-label classification problem**, while the lossy pre-processing pipeline, could be addressed by creating a model that feeds flows and sessions into an **Auto-Encoder** without cropping the **packet payload** at all.

ACKNOWLEDGEMENT

This research has been supported by the European Union's Horizon 2020 Research and Innovation Programme Analysis Software Scheme of Uniform Statistical Sampling, Audit and Defence Processes (SANCUS) under the Grant Agreement number 952672.

CONTACT INFORMATION

Dr. Antonios Lalas, Centre for Research and Technologies Hellas, lalas@iti.gr