5G PFCP Intrusion Detection Dataset – ReadME File

K3Y Ltd - https://k3ylabs.com/

Authors: George Amponis, Panagiotis Radoglou-Grammatikis, George Nakas, Maria Zevgara, Sofia Giannakidou, Savvas Ouzounidis, George Kakamoukas,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).

1. Introduction

The advancements in the field of telecommunications have resulted in an increasing demand for robust, high-speed, and secure connections between User Equipment (UE) instances and the Data Network (DN). The implementation of the newly defined 3rd Generation Partnership Project 3GPP (3GPP) network architecture in the 5G Core (5GC) represents a significant leap towards fulfilling these demands. This architecture promises faster connectivity, low latency, higher data transfer rates, and improved network reliability. 5GC has been designed to support a wide range of critical Next Generation Internet of Things (NG-IoT) and industrial use cases that require reliable end-to-end communication services. However, this evolution raises severe security issues. In the context of the SANCUS[1] project, a set of cyberattacks were investigated and emulated by K3Y against the Packet Forwarding Control Protocol (PFCP) between the Session Management Function (SMF) and the User Plane Function (UPF). Based on these attacks, an intrusion detection dataset was generated: 5GC PFCP Intrusion Detection Dataset that can support the development of Artificial Intelligence (AI)-powered Intrusion Detection Systems (IDS) that use Machine Learning (ML) and Deep Learning (DL) techniques. The goal of this report is to describe this dataset.

[1] https://sancus-project.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).

2. Instructions

The 5GC PFCP Intrusion Detection Dataset was implemented following relevant methodological frameworks, including eleven features: (a) Complete Network Configuration, (b) Complete Traffic, (c) Labelled Dataset, (d) Complete Interaction, (e) Complete Capture, (f) Available Protocols, (g) Attack Diversity, (h) Heterogeneity, (i) Feature Set and (j) Metadata.

A 5GC architecture was emulated, including the Network Slice Selection Function (NSSF), the Network Exposure Function (NEF), the Network Repository Function (NRF), the Policy Control Function (PCF), the User Data Management (UDM), the Access and Mobility Management Function (AF), the Authentication Server Function (AUSF), the Access Management Function (AMF), SMF, and UPF, in addition to a virtualised UE device, a virtualised gNodeB (gNB), and a cyberattacker impersonating a maliciously instantiated SMF. In particular, the following cyberettacks were performed:

- On Wednesday, October 05, 2022, the **PFCP Session Establishment DoS Attack** was implemented for 4 hours.
- On Thursday, October 13, 2022, the **PFCP Session Deletion DoS Attack** was implemented for four hours.
- On Tuesday, November 01, 2022, the **PFCP Session Modification DoS Attack (DROP Apply Action Field Flags)** was implemented for 4 hours.
- On Tuesday, November 22, 2022, the **PFCP Session Modification DoS Attack (DUPL Apply Action Field Flag)** was implemented for 4 hours.

The previous PFCP-related cyberattacks were executed, utilising penetration testing tools, such as `Scapy`[2]. For each attack, a relevant folder is provided, including the network traffic and the network flow statistics for each entity. In particular, for each cyberattack, a folder is given, providing (a) the pcap files for each entity, (b) the Transmission Control Protocol (TCP)/ Internet Protocol (IP) network flow statistics for 120 seconds in a Comma-Separated Values (CSV) format and (c) the PFCP flow statistics for each entity (using different timeout values in terms of second (such as 45, 60, 75, 90, 120 and 240 seconds)). The TCP/IP network flow statistics were produced by using the `CICFlowMeter`[3], while the PFCP flow statistics were generated based on a `Custom PFCP Flow Generator`, taking full advantage of `Scapy`.

[2] *https://scapy.net/*
[3] *https://github.com/ahlashkari/CICFlowMeter*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).

3. Dataset Structure

The dataset consists of the following files.

- **Balanced PFCP APP Layer.7z**: It includes the balanced CSV files from `CICFlowMeter` that may be used to train ML and DL algorithms. Each folder includes a different sub-folder for the corresponding flow timeout values used by the `Custom PFCP Flow Generator`.
- **Balanced TCP-IP Layer.7z**: It includes the balanced CSV files from the `Custom PFCP Flow Generator` that may be used to train ML and DL algorithms. Each folder includes a different sub-folder for the corresponding flow timeout values used by CICFlowMeter.
- **PFCP Session Deletion DoS Attack.7z**: It includes the pcap files and CSV files related to the PFCP Session Deletion Denial of Service (DoS) Attack.
- **PFCP Session Establishment DoS Attack.7z**: It includes the pcap files and CSV files related to the PFCP Session Establishment Flood DoS Attack.
- **PFCP Session Modification DoS Attack.7z**: It includes the pcap files and CSV files related to the PFCP Session Modification DoS Attack.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).

4. Testbed and PFCP Attacks

Figure 1 shows the testbed created for generating this dataset. It is composed of twelve dockerised 5G network functions that emulated, utilising `Open5GS`[4] and `UERANSIM`[5]. Moreover, there is another element serving as the attacker's entry point (as a malicious insider) to the virtualised infrastructure, namely an SMF instance networked in parallel to the original network function. As such, the attacker acts as a malicious insider, executing the cyberattacks against the aforementioned network functions. Finally, the network traffic data of each entity/device was captured through `Tshark`[6] individually for each network function and radio element, respectively.

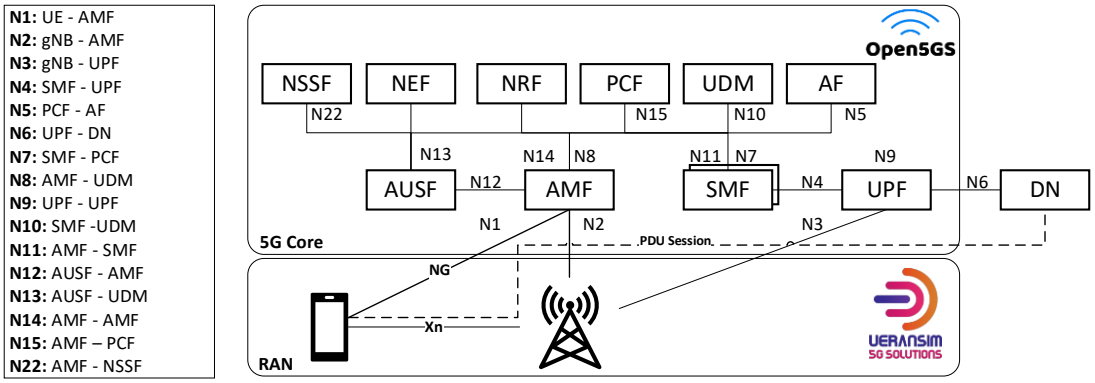

Figure 1: Testbed used for the generation of the PFCP Intrusion Detection Dataset

The description of the PFCP attacks is given in Table 1.

Table 1: Description of the PFCP attacks

PFCP Cyberattack	Description
PFCP Session Establishment Flood Attack	The goal of this attack is the exhaustion of the UPF's resources to handle legitimate Session Establishment Requests and Heartbeat Requests. This will potentially hinder the capability of the 5GC to successfully formulate new Protocol Data Unit (PDU) sessions between clients and DN. This attack is implemented on the N4 interface, and the impact can be observed in the intermediate interfaces. The Session ID (SEID) is randomized for each session establishment request.
PFCP Session Deletion Flood DoS Attack	The goal of this attack is to disassociate a targeted UE from the DN. More specifically, the script targets the PDU sessions between the clients and the DN in a such manner that does not disassociate the UE from the 5G Radio Access Network (RAN) or the Core network but rather only cuts them off the DN. This attack is implemented on the N4 interface, and the impact can be observed in the N6 interface. The only way to re-associate an affected UE is to re-initiate the following procedure: the affected UE can either re-start its session or enter the range of another gNb, at which event a new SEID will be associated with the UE's PDU session, and the attack's effect will be nullified.

[4] *https://open5gs.org/*
[5] *https://github.com/aligungr/UERANSIM*
[6] *https://www.wireshark.org/docs/man-pages/tshark.html*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).

PFCP Session Modification Flood attack (DROP Apply Action Field Flags):	The goal of this attack is to discard packet handling rules for a specific session, thus disassociating a targeted UE from the DN. If the rules are changed successfully, the Forwarding Action Rule (FAR) rules containing the TEID and IP address of the base station are deleted on the UPF. As a result, the GPRS Tunneling Protocol (GTP) tunnel for the subscriber's downlink data is cut off, depriving the subscriber from Internet. The GPRS Tunneling Protocol - User Plane (GTP-U) tunnel can be subsequently restored by sending the re-quired data to the UPF. As with the other PFCP-based attacks, the script targets the PDU sessions between the clients and the DN in a such manner that does not dissasociate the UE from the 5G RAN or the Core network, but rather only cuts them off the DN. This attack is implemented on the N4 interface, and the impact can be observed in the N6 interface.
PFCP Session Modification Flood attack (DUPL Apply Action Field Flag)	The goal of this attack is to use the DUPL flag in the Apply Action field to force the UPF to duplicate the rules for the session, creating multiple paths for the same data from a single source. This may cause undefined behaviour in the N6 interface and/or cause traffic to be duplicated upon transmission towards the DN. Moreover, this attack can be part of a greater scheme geared towards performing a Distributed DoS (DDoS) attack against hosts located in the DN, while also exhausting the UPF's resources to forward outgoing packets to hosts outside the 5GC. By multiplying the number of packets transmitted per active user, a malicious entity can generate a near-passive attack vector which can be easily scaled to affect the traffic of numerous subscribers and thus exponentially exhaust the packet handling resources of the UPF.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).

5. Feuatures

The TCP/IP network flow statistics generated by `CICFlowMeter` are summarised below. It is worth mentioning that the TCP/IP network flows and their statistics generated by `CICFlowMeter` are labelled based on PFCP attacks described above, thus allowing the training of ML/DL models.

Table 2: CICFlowMeter TCP/IP Network Flow Statistics - Features

Feature	Description
Flow ID	ID of the flow
Src IP	Source IP address
Src Port	Source TCP/UDP port
Dst IP	Destination IP address
Dst Port	Destination TCP/UDP port
Protocol	Protocol related to the flow
Timestamp	Flow timestamp
Flow Duration	Duration of the flow in Microseconds
Tot Fwd Pkts	Total packets in forward direction
Tot Bwd Pkts	Total packets in backward direction
TotLen Fwd Pkts	Total size of packets in forward direction
TotLen Bwd Pkts	Total size of packets in backward direction
Fwd Pkt Len Max	Maximum size of packet in forward direction
Fwd Pkt Len Min	Minimum size of packet in forward direction
Fwd Pkt Len Mean	Mean size of packet in forward direction
Fwd Pkt Len Std	Standard deviation size of packet in forward direction
Bwd Pkt Len Max	Maximum size of packet in backward direction
Bwd Pkt Len Min	Minimum size of packet in backward direction
Bwd Pkt Len Mean	Mean size of packet in backward direction
Bwd Pkt Len Std	Standard deviation size of packet in backward direction
Flow Byts/s	Number of flow bytes per second
Flow Pkts/s	Number of flow packets per second
Flow IAT Mean	Mean time between two packets sent in the flow
Flow IAT Std	Standard deviation time between two packets sent in the flow
Flow IAT Max	Maximum time between two packets sent in the flow
Flow IAT Min	Minimum time between two packets sent in the flow
Fwd IAT Tot	Total time between two packets sent in the forward direction
Fwd IAT Mean	Mean time between two packets sent in the forward direction
Fwd IAT Std	Standard deviation time between two packets sent in the forward direction
Fwd IAT Max	Maximum time between two packets sent in the forward direction
Fwd IAT Min	Minimum time between two packets sent in the forward direction
Bwd IAT Tot	Total time between two packets sent in the backward direction
Bwd IAT Mean	Mean time between two packets sent in the backward direction

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).

Bwd IAT Std	Standard deviation time between two packets sent in the backward direction
Bwd IAT Max	Maximum time between two packets sent in the backward direction
Bwd IAT Min	Minimum time between two packets sent in the backward direction
Fwd PSH Flags	Number of Forward PSH flags
Bwd PSH Flags	Number of Backward PSH flags
Fwd URG Flags	Number of Forward URG flags
Bwd URG Flags	Number of Backward URG flags
Fwd Header Len	Length of Forward header
Bwd Header Len	Length of Backward header
Fwd Pkts/s	Number of Forward packets per second
Bwd Pkts/s	Number of Backward packets per second
Pkt Len Min	Minimum packet length
Pkt Len Max	Maximum packet length
Pkt Len Mean	Mean packet length
Pkt Len Std	Standard deviation of packet length
Pkt Len Var	Variance of packet length
FIN Flag Cnt	Number of FIN flags
SYN Flag Cnt	Number of SYN flags
RST Flag Cnt	Number of RST flags
PSH Flag Cnt	Number of PSH flags
ACK Flag Cnt	Number of ACK flags
URG Flag Cnt	Number of URG flags
CWE Flag Count	Number of CWE flags
ECE Flag Cnt	Number of ECE flags
Down/Up Ratio	Down/Up ratio
Pkt Size Avg	Average packet size
Fwd Seg Size Avg	Average Forward segment size
Bwd Seg Size Avg	Average Backward segment size
Fwd Byts/b Avg	Average Forward bytes per bit
Fwd Pkts/b Avg	Average Forward packets per bit
Fwd Blk Rate Avg	Average Forward block rate
Bwd Byts/b Avg	Average Backward bytes per bit
Bwd Pkts/b Avg	Average Backward packets per bit
Bwd Blk Rate Avg	Average Backward block rate
Subflow Fwd Pkts	Number of Forward subflow packets
Subflow Fwd Byts	Number of Forward subflow bytes
Subflow Bwd Pkts	Number of Backward subflow packets
Subflow Bwd Byts	Number of Backward subflow bytes
Init Fwd Win Byts	Initial Forward window bytes
Init Bwd Win Byts	Initial Backward window bytes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).

Fwd Act Data Pkts	Number of Forward active data packets
Fwd Seg Size Min	Minimum Forward segment size
Active Mean	Mean active time
Active Std	Standard deviation of active time
Active Max	Maximum active time
Active Min	Minimum active time
Idle Mean	Mean idle time
Idle Std	Standard deviation of idle time
Idle Max	Maximum idle time
Idle Min	Minimum idle time
Label	Label

On the other hand, the PFCP flow statistics generated by the `Custom PFCP Python Parser` are summarised below.

Table 3: PFCP Network Flow Statistics – Features

Feature	Description
flow ID	Flow identifier
source IP	Source IP address
destination IP	Destination IP address
source port	Source port number
destination port	Destination port number
protocol	Network layer protocol
duration	Length of time flow was active
fwd_packets	Number of forward packets
bwd_packets	Number of backward packets
PFCPHeartbeatRequest_counter	Number of PFCP Heartbeat Request messages
PFCPHeartbeatResponse_counter	Number of PFCP Heartbeat Response messages
PFCPPFDManagementRequest_counter	Number of PFCP PFD Management Request messages
PFCPPFDManagementResponse_counter	Number of PFCP PFD Management Response messages
PFCPAssociationSetupRequest_counter	Number of PFCP Association Setup Request messages
PFCPAssociationSetupResponse_counter	Number of PFCP Association Setup Response messages
PFCPAssociationUpdateRequest_counter	Number of PFCP Association Update Request messages
PFCPAssociationUpdateResponse_counter	Number of PFCP Association Update Response messages
PFCPAssociationReleaseRequest_counter	Number of PFCP Association Release Request messages

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).

PFCPAssociationReleaseResponse_counter	Number of PFCP Association Release Response messages
PFCPVersionNotSupportedResponse_counter	Number of PFCP Version Not Supported Response messages
PFCPNodeReportRequest_counter	Number of PFCP Node Report Request messages
PFCPNodeReportResponse_counter	Number of PFCP Node Report Response messages
PFCPSessionSetDeletionRequest_counter	Number of PFCP Session Set Deletion Request messages
PFCPSessionSetDeletionResponse_counter	Number of PFCP Session Set Deletion Response messages
PFCPSessionEstablishmentRequest_counter	Number of PFCP Session Establishment Request messages
PFCPSessionEstablishmentResponse_counter	Number of PFCP Session Establishment Response messages
PFCPSessionModificationRequest_counter	Number of PFCP Session Modification Request messages
PFCPSessionModificationResponse_counter	Number of PFCP Session Modification Response messages
PFCPSessionDeletionRequest_counter	Number of PFCP Session Deletion Request messages
PFCPSessionDeletionResponse_counter	Number of PFCP Session Deletion Response messages
PFCPSessionReportRequest_counter	Number of PFCP Session Report Request messages
PFCPSessionReportResponse_counter	Number of PFCP Session Report Response messages
Downlink_counter	Number of downlink packets
Uplink_counter	Number of uplink packets
Bidirectional_traffic_counter	Number of bidirectional traffic packets
Label	Flow label (e.g. benign or malicious)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).

6. Balanced Versions

Two balanced versions of the dataset have been created for the flows generated by `CicFlowMeter` and the `Custom PFCP Python Parser` to support the development of AI-powered detection solutions. The balanced files are involved in the following folders.

- **Balanced_PFCP_APP_Layer**: This folder contains the PFCP flow statistic generated by the `Custom Python PFCP Generator`.
- **Balanced_TCP-IP_Layer**: This folders contains the TCP-IP flow statistics generated by `CICFlowMeter`.

Each folder contains a set of subfolders for each flow timeout, as given below.

- **15-sec-CSV**: The flow timeout was set to 15 seconds
- **20-sec-CSV**: The flow timeout was set to 15 seconds
- **60-sec-CSV**: The flow timeout was set to 60 seconds
- **120-sec-CSV**: The flow timeout was set to 120 seconds
- **240-sec-CSV**: The flow timeout was set to 240 seconds

Each version is balanced. Therefore, they contain an equal number of classes.

The five classes of the dataset and their respective labels are provided. The classes are given in the following table.

Table 4: Dataset labels

Class	Label
Normal flow	"Normal"
PFCP Session Establishment Flood attack flow	"Mal_Estab"
PFCP Session Deletion Flood attack flow	"Mal_Del"
PFCP Session Modification Flood attack (DROP Apply Action Field Flags) flow	"Mal_Mod"
PFCP Session Modification Flood attack (DUPL Apply Action Field Flag) flow	"Mal_Mod2"

For each flow timeout, there are two sub-subfolders:

- **Training**: It includes the data samples that can be used for the training procedure of an AI model.
- **Testing**: It includes the data samples that can be used for the testing procedure of an AI model.

Each of these sub-subfolders contains a .csv file named "`Training_X.csv`" and "`Testing_X.csv`", where *X* is the flow timeout value, for the training and the testing of the AI models, respectively. The split ratio is: 70% - 30%, for Training-Testing respectively. In addition, the splitting is stratified, meaning that the same percentage of samples of each class are present in the Training and Testing.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).

The number of flows for each flow timeout value for the TCP/IP network flow statistics generated by CICFlowMeter are given below.

Table 5: Number of the TCP/IP flows (generated by `CICFlowMeter`) for the different flow timeout values in the balanced files.

Training					
Timeout	**Normal**	**Mal_Estab**	**Mal_Del**	**Mal_Mod**	**Mal_Mod2**
15s	1439	1440	1440	1440	1440
20s	1439	1440	1440	1440	1440
60s	485	485	485	485	485
120s	260	261	260	260	261
240s	133	134	133	134	134
Testing					
Timeout	**Normal**	**Mal_Estab**	**Mal_Del**	**Mal_Mod**	**Mal_Mod2**
15s	618	617	617	617	617
20s	618	617	617	617	617
60s	208	208	208	208	208
120s	112	111	112	112	111
240s	58	57	58	57	57

The number of flows for each flow timeout value for the TCP/IP network flow statistics generated by `CICFlowMeter` are given below.

Table 6: Number of the PFCP flows (generated by `Custom PFCP Flow Generator`) for the different flow timeout values in the balanced files.

Training					
Timeout	**Normal**	**Mal_Estab**	**Mal_Del**	**Mal_Mod**	**Mal_Mod2**
15s	1103	1104	1104	1104	1104
20s	666	667	666	666	667
60s	222	223	222	223	223
120s	110	111	110	111	111
240s	68	69	68	69	69
Testing					
Timeout	**Normal**	**Mal_Estab**	**Mal_Del**	**Mal_Mod**	**Mal_Mod2**
15s	474	473	473	473	473
20s	286	285	286	286	285
60s	96	95	96	95	95
120s	48	47	48	47	47
240s	30	29	30	29	29

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).

7. Citation

The users of this dataset are kindly asked to cite the following paper(s) as follows.

G. Amponis, P. Radoglou-Grammatikis, T. Lagkas, W. Mallouli, A. Cavalli, D. Klonidis, E. Markakis, and P. Sarigiannidis, "Threatening the 5G core via PFCP DOS attacks: The case of blocking UAV Communications", EURASIP Journal on Wireless Communications and Networking, vol. 2022, no. 1, 2022, doi: 10.1186/s13638-022-02204-5.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952672 (SANCUS).