

# An ELEGANT dataset with Denial of Service and Man in The Middle attacks

Bruno Sousa  
CISUC, DEI  
University of Coimbra  
Coimbra, Portugal  
bmsousa@dei.uc.pt

Tiago Cruz  
CISUC, DEI  
University of Coimbra  
Coimbra, Portugal  
tjcruz@dei.uc.pt

Miguel Arieiro  
CISUC, DEI  
University of Coimbra  
Coimbra, Portugal  
marieiro@student.dei.uc.pt

Vasco Pereira  
CISUC, DEI  
University of Coimbra  
Coimbra, Portugal  
vasco@dei.uc.pt

**Abstract**—This document describes a dataset with diverse types of Denial of Service (DoS) attacks and Man-in-the-Middle (MiTM) attacks. The data is available online and reachable via the DOI 10.21227/mewp-g646. This document describes the data collection process and provides useful information on how such data can be employed to devise models for cybersecurity in critical infrastructures using Programmable Logic Controllers (PLCs)

**Index Terms**—Denial of Service, Man in the Middle, PLC, cybersecurity, dataset

## I. INTRODUCTION

Modern automation technologies have become pervasive [1], playing a crucial part in ensuring the delivery and availability of several essential services. As a result, operators and service utilities are often compelled by stakeholders, governmental bodies, as well as regulatory, standardisation and steering organisations, to improve service quality, security and reliability. In this regard, the constant monitoring of control elements (i.e., Programmable Logic Controllers - PLCs) is crucial for management and also to detect anomalous behaviour in Industrial Automation Control Systems (IACS).

PLCs, running protocols like Modbus/TCP [2], are often targeted by diverse types of attacks. These may include amplification or network flooding attempts, for Denial of Service (DoS) [3] purposes, as well as more sophisticated techniques, such as Man in the Middle (MiTM) attacks, using techniques like Address Resolution Protocol (ARP) poisoning to tamper with process information (e.g. water pumps, water levels, temperature sensors, turbine speed sensors, etc) and/or cause loss of visibility. As an example of the latter case, an attacker might leverage MiTM techniques to manipulate temperature sensor measurements: while a sensor may report a real value of 120°C, an attacker may be able to modify inflight telemetry to deceive PLCs and other Human Machine Interface (HMI) elements to assume and report a temperature reading of 40°C.

This paper documents one of the outcomes achieved in the ELEGANT project, funded by the 7<sup>th</sup> Open call of Fed4Fire+ for large experiments<sup>1</sup>, namely the dataset with DDoS and MiTM attacks that is publicly available at the IEEE DataPort

<sup>1</sup>The data collected resulted from the funding of FedFire+ Open calls - large experiments.

<sup>1</sup>7th OpenCall Fed4Fire+ Large Experiments.

repository (DOI: 10.21227/mewp-g646). The dataset includes flooding and amplification DDoS attacks performed from single and multiple nodes, as well as ARP-poisoning based MiTM attacks.

## II. COLLECTION PROCESS

### A. Overall architecture

The base architecture that was deployed in Fed4Fire+ testbeds (virtual Wall2 and Grid5000) and used to collect the data available in the dataset is pictured in Figure 1.

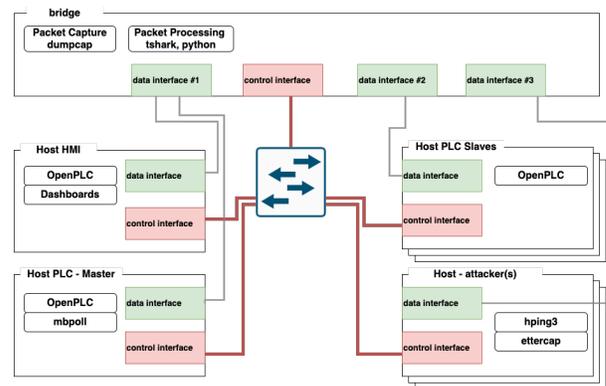


Fig. 1. Overall architecture with PLCs

All the components have a data and a control interface. All the relevant information, namely the Modbus TCP [2] data packets exchanged in the scope of the process control tasks, were collected in the data interfaces. The capture was performed in the bridge node using the dumpcap tool [4].

The PLC master node queries other PLC nodes regarding the information of sensors stored in specific registers. PLC slaves have sensors information which is stored in the holding registers as per the PLC internal mechanism. It should be noticed that multiple PLCs exist in the testbed. All the PLC are based on the OPenPLC v3 version [5].

### B. PLC Modbus settings

Regarding the reference topology, there are both horizontal (PLC-PLC) and vertical (PLC-HMI) communication patterns. PLCA1 has multiple configured slaves, which are polled in

100ms intervals. Such polling can be reduced in real-world deployments, for instance to values around 40ms, leading to rates around 25pkts/s [6]. To determine the rates in the attacks, we consider 25pkts/s as the reference rate for the Modbus TCP polling process.

PLC master queries the diverse PLCs (e.g. PLC slave) which implement the logic for querying information from sensors, also performing actuation tasks based on such information. Figure 2 illustrates the functionality of the program running at the PLCs. The underlying logic considers a process for water level control in a tank, using a level sensor and a water pump (labelled as SWITCH in Figure. 2) which is activated if a certain threshold is achieved.

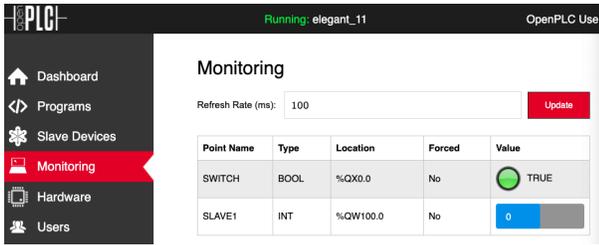


Fig. 2. ELEGANT Water level program variables

### C. Denial of Service attacks

The Denial of Service (DoS) attacks have been performed considering multiple configurations. Two traffic rate classes were considered for the inter-arrival packet times. Rate 1 includes packet interval rates in a ratio of 10 times bigger the normal rate of the Modbus TCP polling process ( $10 \times normalRate$ ). Rate 2 assumes the maximum flood that can be performed with the attack node ( $--flood$  option), as summarized in Table I.

TABLE I  
DoS ATTACK RATE SETTINGS

Rate	Interval (ms)	N. packets
1	400	2500
2	< 1	Max. supported by node

Table II summarizes the tests that were performed, illustrating the time periods (CET timezone). Each test was performed using the hping3 tool [7] with the ability to randomize source node IP ( $--rand-source$  option). The DoS targeted the PLCA1, which acts as a master, polling information from other devices (other PLCs, such as PLCA2).

The DoS attack types encompass flooding attempts, which target the modbus TCP port (502) with a packet size of 120 bytes (size considered as per the size of the traffic in the polling process of the program running in the PLCs, recall subsection II-B). DoS-based amplification attacks rely on UDP traffic for the same port, but with a size of 60 bytes.

The beginning and end of the attack timeframes are also documented on the respective pcap files, collected with the

dumpcap tool, which was configured to dump network traffic in a ring buffer with file sizes between 10 and 20MB.

The start of the attack is present in the pcap files, as illustrated in Figure 3. In particular, a UDP data packet with a length of 23 bytes is sent to port 503, on the destination PLC. The data contained in the data field of the packet contains the string "ATTACK\_START\_DDOS".

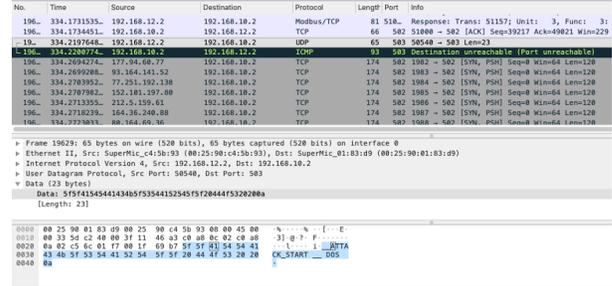


Fig. 3. Start label of the attack in dataset

The end of the attack, as illustrated in Figure 4, consists of a UDP message sent to port 502, with a size of around 90 bytes. The data field contains information of the attack that was performed. For instance, as illustrated in Figure 4, one can observe the label "ATTACK\_END\_DOS hping3".

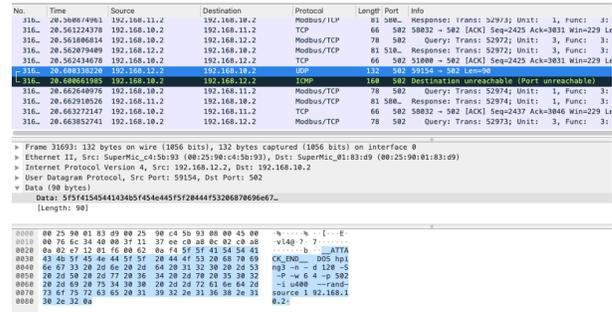


Fig. 4. End label of the attack in dataset

### D. Man in the middle attacks

The technique selected to deploy the MiTM attacks executed within the scope of the ELEGANT dataset extraction process is based on the ARP poisoning technique. ARP is a stateless protocol, designed in such a way what it implicitly relies on unprotected Layer 2/3 mechanisms to provide a dynamic association mechanism that binds MAC to IP addresses within the same link-layer domain.

By design, ARP will cache any replies, whether solicited or not, overwriting non-expired entries when new replies are received. Due to the lack of authentication and authorisation mechanisms, ARP constitutes a weak spot that can be exploited to implement MiTM attacks [8] against IPv4 network environments.

ARP-poisoning MiTM attacks can be deployed for different purposes [9]. For instance, they may provide the means for an attacker to discreetly scout a target infrastructure, capturing

TABLE II  
DOS ATTACKS

Node (S)ingle, (M)ultiple	(S)ingle, type	rate	Date	Hour interval (CET)	Files/ Timestamp
S	Flood	1	2021/03/13	19:44-19:47	00001-00008
S	Flood	1	2021/03/13	19:55-19:58	00008-00016
S	Flood	2	2021/03/13	20:02-20:05	00016-00249
S	Flood	2	2021/03/13	20:10-20:13	00249-00485
S	Amp	-	2021/03/13	20:20-20:24	00485-00661
S	Amp	-	2021/03/13	20:29-20:32	00661-00841
M	Flood	1	2021/03/14	01:02-01:05	20210314005755- 20210314010515
M	Flood	1	2021/03/14	01:37-01:40	20210314013327- 20210314014020
M	Flood	2	2021/03/14	01:12-01:15	20210314010944- 20210314011526
M	Flood	2	2021/03/14	01:19-01:22	20210314011526- 20210314012226
M	Amp	-	2021/03/14	01:44-01:47	20210314014020- 20210314014821
M	Amp	-	2021/03/14	01:51-01:55	20210314014821- 20210314015502

network traffic for analysis as part of the early planning or attack preparation stages. But these attacks can also be useful in offensive roles, allowing attackers to hijack Modbus TCP connections and change/mask relevant information, effectively blinding the operator and disrupting process operation.

Fig. 5. Start label of the MiTM attack in dataset

The Man in the Middle attacks have been performed in two flavors, as summarized in Table III, one considering only the A-ARP poisoning, and another considering the F-full chain of the attack, which modifies data regarding Modbus/TCP protocol. The attacks have been performed to last around 10 minutes, with the exception of the test in the full chain in run 2, which lasts around 6 minutes. A long period with data without MITM attacks is also provided (labelled as no attack in Table III). The attacks are identified in the traces with the protocol UDP at destination port 502 and with the string “\_MiTM\_ATTACK\_START\_” in the data field, as illustrated in Figure. 5.

The end of attacks are identified with packets sent to destination port 502 and with the data field with a string regarding the pattern “\_MiTM\_ATTACK\_END\_”.

The MiTM attack is performed with the ettercap tool [10] in text mode with the option -M arp. The full chain of the attack with packet manipulation is enabled with a filter (option -F name\_filter) to manipulate the information in the Modbus reply messages regarding the values of the records.

Fig. 6. End label of the MiTM attack in dataset

TABLE III  
MITM WITH (A)RP POISONING AND (F)ULL CHAIN ATTACKS

type (A),(F)	run	Date	Hour interval (CET)	Files/ Timestamp
A	1	2021/03/18	03:40-03:50	20210318034103
A	2	2021/03/18	08:50-09:02	20210318084952
F	1	2021/03/18	09:52-10:02	20210318095210
F	2	2021/03/18	10:04-10:10	20210318100414
F	3	2021/03/18	10:16-10:26	20210318101444
no attack		2021/03/18	03:53-08:46	20210318035345- 20210318080920

### III. DATASET STRUCTURE AND FILES

The files available in the dataset are encoded using the pcap format, which stores information in a structured way. The structure considers a global header [11] with the following structure:

```
typedef struct pcap_hdr_s {
    guint32 magic_number; /* magic number */
    guint16 version_major; /* major ver. num. */
    guint16 version_minor; /* minor ver. num. */
    gint32 thiszone; /* GMT to local */
    guint32 sigfigs; /* timestamp accuracy */
    guint32 snaplen; /* max length */
    guint32 network; /* data link type */
} pcap_hdr_t;
```

More information regarding the meaning of the diverse fields is available at [11].

In addition, each record (capture packet) is identified by an header, with the following information:

```
typedef struct pcaprec_hdr_s {
    guint32 ts_sec; /* timestamp seconds */
    guint32 ts_usec; /* timestamp microsec.*/
    guint32 incl_len; /* number of octets */
    guint32 orig_len; /* length of packet */
} pcaprec_hdr_t;
```

A CSV file is also available to illustrate the structure of the collected information, see Figure 7.

No.	Time	Source	Destination	Protocol	Length	Info
1	0.00000000	192.168.10.2	192.168.11.2	Modbus/TCP	78	Query: Trans: 47891; Unit: 1; Func: 3: Read Holding Registers
2	0.000245508	192.168.11.2	192.168.10.2	Modbus/TCP	81	Response: Trans: 47891; Unit: 1; Func: 3: Read Holding Registers
3	0.000626239	192.168.10.2	192.168.11.2	TCP	66	58032 > 502 [ACK] Seq=13 Ack=16 Win=229 Len=0 TSval=340602092 TSecr=1065398478
4	0.001063554	192.168.10.2	192.168.12.2	Modbus/TCP	78	Query: Trans: 47890; Unit: 3; Func: 3: Read Holding Registers
5	0.001314165	192.168.12.2	192.168.10.2	Modbus/TCP	81	Response: Trans: 47890; Unit: 3; Func: 3: Read Holding Registers
6	0.001662244	192.168.10.2	192.168.12.2	TCP	66	51000 > 502 [ACK] Seq=13 Ack=16 Win=229 Len=0 TSval=1545815710 TSecr=2648693038
7	0.101907108	192.168.10.2	192.168.11.2	Modbus/TCP	78	Query: Trans: 47892; Unit: 1; Func: 3: Read Holding Registers
8	0.105483842	192.168.11.2	192.168.10.2	Modbus/TCP	81	Response: Trans: 47892; Unit: 1; Func: 3: Read Holding Registers

Fig. 7. Example of collected data and available in the CSV file

The dataset contains several files regarding the DoS attacks, which are organized as follows:

- 1) *cap\_001\_2021\_DoS\_example.csv*- Sample CV file with data from DoS attack in single more and date 1.
- 2) *single\_flood\_rate\_1\_t1.zip*- DoS flood attack in single node with rate 1 and run 1.
- 3) *single\_flood\_rate\_1\_t2.zip*- DoS flood attack in single node with rate 1 and run 2.
- 4) *single\_flood\_rate\_2\_t1.zip*- DoS flood attack in single node with rate 2 and run 1.
- 5) *single\_flood\_rate\_2\_t2.zip*- DoS flood attack in single node with rate 2 and run 2.
- 6) *single\_amp\_t1.zip*- DoS amplification attack in single node and run 1.
- 7) *single\_amp\_t2.zip*- DoS amplification attack in single node and run 2.
- 8) *multiple\_flood\_0100\_0125.tar.gz*- DoS flood attack in multiple nodes with rate 1 with run 1, and rate 2 with runs 1 and 2.
- 9) *multiple\_amp\_0130\_0200.tar.gz*- DoS flood attack in multiple nodes with rate 1 with run 2, and DoS amplification attack in multiple nodes with run 1 and 2.

The dataset contains several files regarding the DoS attacks, which are organized as follows:

- 1) *MiTM\_ARP\_Poisoning.tar.gz*- MiTM attack with ARP poisoning with run 1 and run 2.
- 2) *MiTM\_Full\_Chain.tar.gz*- MiTM attack with Full Chain, including ARP poisoning attack and data manipulation in run 1 and run 2.
- 3) *normal\_PLC\_traffic.tar.gz*- Regular ModBus/TCP traffic without attacks.

#### IV. CONCLUSION

We hope this dataset may help others working in the research and development of secure solutions for Critical

Infrastructures.

Also, the information herein contained will be updated with more details and developments regarding the achievements and results of the ELEGANT project.

#### REFERENCES

- [1] G. Pretticco, M. Flammini, N. Andreadou, S. Vitiello, G. Fulli, and M. Masera, "Distribution system operators observatory 2018," , Publications Office of the European Union.
- [2] M. Organization, "Modbus messaging on tcp/ip implementation guide v1.0b," October 2006. [Online]. Available: [https://modbus.org/docs/Modbus\\_Messaging\\_Implementation\\_Guide\\_V1\\_0b.pdf](https://modbus.org/docs/Modbus_Messaging_Implementation_Guide_V1_0b.pdf)
- [3] A. Huseinovic, S. Mrdovic, K. Bicakci, and S. Uludag, "A Survey of Denial-of-Service Attacks and Solutions in the Smart Grid," *IEEE Access*, vol. 8, pp. 177 447–177 470, 2020.
- [4] Wireshark, "Wireshark user's guide v3.5.0." [Online]. Available: [https://www.wireshark.org/docs/wsug\\_html\\_chunked/](https://www.wireshark.org/docs/wsug_html_chunked/)
- [5] T. R. Alves, M. Buratto, F. M. de Souza, and T. V. Rodrigues, "Openplc: An open source alternative to automation," in *IEEE Global Humanitarian Technology Conference (GHTC 2014)*, 2014, pp. 585–589.
- [6] M. Niedermaier, F. Fischer, D. Merli, and G. Sigl, "Network Scanning and Mapping for IIoT Edge Node Device Security," in *2019 International Conference on Applied Electronics (AE)*. IEEE, sep 2019, pp. 1–6.
- [7] antirez, "hping3 network tool." [Online]. Available: <https://github.com/antirez/hping>
- [8] C. Foglietta, D. Masucci, C. Palazzo, R. Santini, S. Panzieri, L. Rosa, T. Cruz, and L. Lev, "From detecting cyber-attacks to mitigating risk within a hybrid environment," *IEEE Systems Journal*, vol. 13, no. 1, pp. 424–435, 2019.
- [9] L. Rosa, T. Cruz, P. Simões, E. Monteiro, and L. Lev, "Attacking scada systems: A practical perspective," in *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, 2017, pp. 741–746.
- [10] E. project, "Ettercap home." [Online]. Available: <https://www.ettercap-project.org/>
- [11] [Online]. Available: <https://gitlab.com/wireshark/wireshark/-/wikis/Development/LibpcapFileFormat>