Cyber Range Open-Source para Estação de Tratamento de Efluentes: Demonstração de Ataques MitM/FDI via Modbus TCP e Defesa Ativa na Camada de Enlace

Letícia dos Santos Matos[1], **João Paulo Fernandes de Cerqueira César**[2], **Otávio de Souza Martins Gomes**[1]

[1]Universidade Federal de Itajubá (UNIFEI)
Itajubá – MG – Brasil

[2]Instituto Federal de Minas Gerais – Campus Ouro Preto (IFMG)
Ouro Preto – MG – Brasil

matosleticia.tec@gmail.com, joaopaulofcc@gmail.com, otavio.gomes@unifei.edu.br

Abstract. *This paper presents a Wastewater Treatment Plant (WWTP) cyber range built with open-source tools: OpenPLC, FUXA SCADA, and a Python-based Digital Twin, orchestrated via Docker over Modbus TCP/IP. The convergence of IT and OT exposes this type of critical infrastructure to cyberattacks with environmental and public health consequences, while reference cyber ranges rely on proprietary hardware. On the proposed platform, Man-in-the-Middle attacks with actuator manipulation and False Data Injection into level registers are reproduced, inducing decoupling between the actual process state and the operator display. As a countermeasure, an Active Defense at the Data Link Layer using bidirectional Gratuitous ARP is proposed, which neutralizes standard ARP spoofing attacks under the tested conditions without requiring cryptography on field devices.*

Resumo. *Este trabalho apresenta um cyber range de Estação de Tratamento de Efluentes (ETE) construído com ferramentas de código aberto: OpenPLC, FUXA SCADA e Gêmeo Digital em Python, orquestrados via Docker sobre Modbus TCP/IP. A convergência entre TI e TO expõe esse tipo de infraestrutura crítica a ataques cibernéticos de consequências ambientais e sanitárias, enquanto os cyber ranges de referência são baseados em hardware proprietário e de alto custo. Sobre a plataforma proposta, reproduzem-se ataques de Man-in-the-Middle com manipulação de atuador e Falsa Injeção de Dados em registradores de nível, induzindo desacoplamento entre o estado real do processo e o valor exibido ao operador. Como contramedida, propõe-se Defesa Ativa na Camada de Enlace por Gratuitous ARP bidirecional, que neutraliza ataques de ARP spoofing padrão nas condições testadas, sem exigir criptografia nos dispositivos de campo.*

1. Introdução

A convergência entre Tecnologia da Informação (TI) e Tecnologia de Operação (TO) expôs Sistemas de Controle Industrial (ICS) a vetores de ameaça antes confinados ao

domínio corporativo. Sistemas historicamente isolados por *air gap* foram integrados a redes IP sem os mecanismos de segurança correspondentes [6]. Beresford [2] demonstrou, em 2011, uma cadeia de exploração contra CLPs Siemens S7 com desvio de autenticação, evidenciando que o Modbus, protocolo adotado pelo OpenPLC neste projeto, foi concebido sem autenticação ou confidencialidade. Em fevereiro de 2021, um atacante remoto acessou o sistema SCADA da ETE de Oldsmar (FL) e elevou a concentração de hidróxido de sódio de 100 ppm para 11.100 ppm, 111 vezes a concentração de operação normal, antes da reversão manual pelo operador [1], expondo que sistemas de saneamento frequentemente operam sem segmentação de rede ou autenticação nos protocolos de campo.

Estações de tratamento de efluentes integram infraestruturas críticas de saneamento; seu comprometimento pode resultar em contaminação de corpos hídricos e violação de padrões sanitários [6]. A análise de segurança em plantas reais impõe riscos inaceitáveis de interrupção de processo [6]. *Cyber ranges* virtuais eliminam essa barreira ao reproduzir o comportamento do processo em ambiente controlado e de custo próximo de zero. Este trabalho descreve um *cyber range* de ETE construído integralmente com ferramentas de código aberto, sobre o qual demonstram-se ataques MitM e FDI via Modbus TCP e uma contramedida operacional na Camada de Enlace.

2. Trabalhos Relacionados

A NIST SP 800-82 [6] cataloga ameaças e contramedidas para SCADA, DCS e CLPs; sua terceira revisão incorpora gêmeos digitais como superfícies emergentes. Beresford [2] demonstrou cadeia de exploração contra Siemens S7 sem proposta de defesa. Este trabalho endereça essa lacuna na Camada de Enlace. Morris e Gao [5] construíram *testbeds* físicos com tanque de água e duto de gás para gerar tráfego Modbus sob 28 classes de ataque, sem virtualização ou análise de contramedidas. O SWaT [4] é o *benchmark* internacional mais impactante para plantas de tratamento de água, com seis estágios em CLPs Allen-Bradley, porém exige hardware proprietário e não incorpora gêmeo digital. Liu, Ning e Reiter [3] formalizaram FDI em redes elétricas com modelo de estimação de estado DC, demonstrando que dados falsos podem contornar detectores de inconsistência; o modelo é teórico e restrito ao domínio elétrico. No contexto nacional, Fritz [9] documenta a automação de uma ETE com CLP Siemens S7-300 sem análise de segurança. Sanchez [8] e Alves et al. [7] documentam MitM contra Modbus TCP mas não propõem contramedida na Camada 2. Propostas como S-ARP [10] exigem infraestrutura de chaves incompatível com CLPs legados.

O presente trabalho substitui hardware proprietário por pilha integralmente *open-source* em Docker, estende o *conceito* de FDI ao domínio Modbus TCP em ETE e documenta experimentalmente a Defesa Ativa por GARP bidirecional sem alteração nos dispositivos de campo.

3. Metodologia

A abordagem experimental estrutura-se em três etapas: construção de um *cyber range* virtualizado, execução de ataques controlados sobre Modbus TCP/IP e avaliação de uma contramedida na Camada de Enlace.

3.1. Arquitetura do Cyber Range

O *cyber range* é composto por um *host* físico e dois contêineres Docker interconectados em rede *bridge* dedicada (Figura 1). **(1) Host físico / OpenPLC Runtime** executa lógica ST conforme IEC 61131-3, gerenciando quatro válvulas (C-01–C-04), duas bombas e um misturador; atuadores expostos em *Coils* via FC01 e sensores em *Holding Registers* via FC03/FC16. O Gêmeo Digital em Python, executado neste mesmo *host*, atualiza o nível $L_i(t)$ de cada tanque por incremento/decremento fixo a cada ciclo de tempo, conforme o estado dos atuadores (bombas e válvulas), escrevendo resultados nos *Holding Registers* `HR 0`–`HR 2` via FC16. **(2) Contêiner FUXA SCADA** realiza *polling* Modbus cíclico na porta TCP 502, lendo *Coils* e *Holding Registers* para atualizar o sinótico (Figura 2) e enviar comandos FC05 ao CLP. **(3) Contêiner Kali Linux**, com flag `--privileged`, executa `arpspoof`, `Scapy` e `NetfilterQueue`. A comunicação ocorre exclusivamente via Modbus TCP/IP sem criptografia ou autenticação [6, 2]. Scripts, imagens Docker e lógica ST utilizados na construção deste *cyber range* estão disponíveis publicamente em repositório para replicação.[1]

[1] `https://anonymous.4open.science/r/Cyber-Range-ETE-3705/README.md`

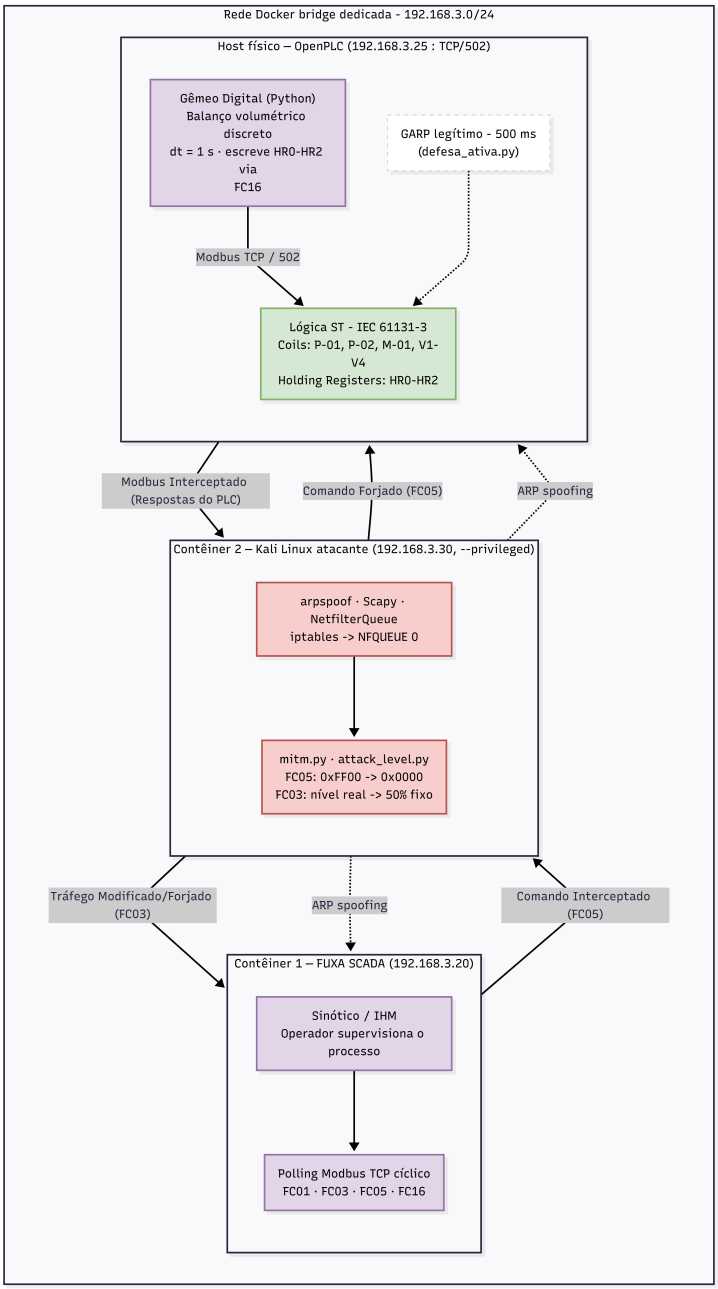

Figura 1. Arquitetura do *cyber range*: host físico (OpenPLC + Gêmeo Digital + `defesa_ativa.py`), Contêiner 1: FUXA SCADA (192.168.3.20) e Contêiner 2: Kali Linux (192.168.3.30). Setas tracejadas: ARP Spoofing; setas sólidas: reescrita FC05/FC03 via NFQUEUE 0.

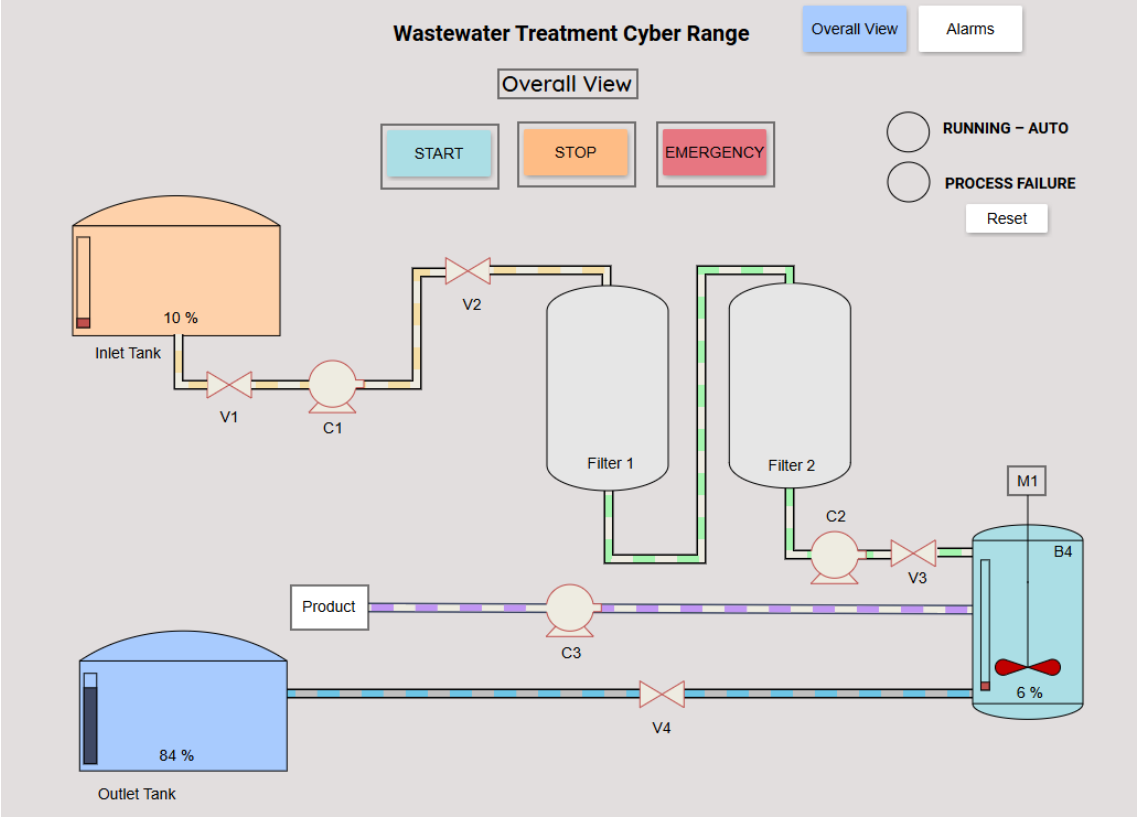

Figura 2. Tela *Overview* do FUXA SCADA: sinótico da ETE com tanques, filtros, bombas C1–C3 e válvulas V1–V4.

3.2. Ataque 1: Man-in-the-Middle com Manipulação de Atuador

ARP Spoofing bidirecional via `arpspoof` posicionou o contêiner Kali como *proxy* transparente entre o contêiner FUXA e o *host* OpenPLC; pacotes foram encaminhados para `NFQUEUE 0` via `iptables`. O script `mitm.py` identificou pacotes FC05 destinados à válvula C-03 (*Coil* `0x0002`), substituiu `0xFF00` (ABRIR) por `0x0000` (FECHAR) e adulterou as respostas do CLP, mantendo o sinótico com ícone de *running* enquanto a válvula permanecia fisicamente fechada. O redirecionamento foi verificado por inspeção das tabelas ARP nos contêineres envolvidos.

3.3. Ataque 2: False Data Injection em Sensor de Nível

O segundo cenário estende o posicionamento MitM para atacar a integridade dos dados de processo. O ataque segue o conceito de FDI de Liu et al. [3]: dados falsos que contornam detectores de inconsistência sem exigir conhecimento do modelo de estimação de estado. Aplicado ao domínio Modbus TCP, a injeção ocorre sobre valores brutos de registradores, sem necessidade de modelo de processo. Após o mesmo posicionamento ARP do Cenário 1, o script `attack_level.py` interceptou as respostas FC03 e substituiu o valor real do *Holding Register* do Tanque de Entrada por 50% fixo. Com o tanque em 100% (transbordamento), o FUXA continuou exibindo 50%, como evidenciado na Listagem 1.

Listing 1. Saída do `attack_level.py`.

```
[*] MitM ativo. Interceptando FC03...
[>] Real (OpenPLC): Nivel T1 = 100 %  (transbordamento)
```

```
[<] Injetado (FUXA): Nivel T1 =  50 %  (falso - estatico)
[!] SCADA: nenhum alarme disparado.
```

Como o OpenPLC lê os registradores escritos pelo Gêmeo Digital e não os devolvidos ao FUXA, o CLP operou sobre o estado real; o *Decoupling* manifestou-se exclusivamente na camada de supervisão.

3.4. Defesa Ativa na Camada 2

A contramedida é um IPS baseado em *Gratuitous ARP* (GARP) bidirecional executado no *host* OpenPLC. Diferentemente de S-ARP [10], que exige infraestrutura de chaves, a abordagem opera integralmente na Camada 2 sem modificar os dispositivos de campo. O script `defesa_ativa.py` resolve dinamicamente os MACs reais do contêiner FUXA e do *gateway* Docker e anuncia os pares IP/MAC corretos a cada 500 ms, sobrescrevendo qualquer entrada forjada pelo atacante. O protocolo de validação usou cinco terminais: T1 (`detect_mitm.py`); T2 (`defesa_ativa.py`, iniciado *antes* do ataque); T3–T4 (`arpspoof` bidirecional); T5 (`iptables DROP` porta 502 e `attack_simple.py`). Este mecanismo materializa, no nível de protocolo, a defesa declarada no resumo: por operar exclusivamente na Camada de Enlace, a contramedida independe do conteúdo das mensagens Modbus TCP trafegadas nas Camadas 4 e 7, o que a torna aplicável a qualquer dispositivo de campo legado sem suporte a criptografia.

4. Resultados e Discussão

4.1. Eficácia dos Ataques MitM e FDI

O envenenamento ARP bidirecional estabeleceu-se imediatamente após a ativação dos Terminais 3 e 4, verificado por inspeção das tabelas ARP. A substituição opera em nível de pacote individual: todo FC05 destinado ao *Coil* `0x0002` com valor `0xFF00` é reescrito para `0x0000` antes de alcançar o OpenPLC: falha silenciosa indistinguível de operação normal para o operador a partir das indicações nativas do SCADA, na ausência de mecanismos adicionais de detecção como IDS por invariantes de processo ou monitoramento de estado físico independente.

Para o FDI, o `attack_level.py` injetou 50% fixo no HR 0: o FUXA manteve leitura constante enquanto o Gêmeo Digital registrou o nível real evoluindo até 100% (transbordamento). A Figura 3 ilustra a evolução temporal do *Decoupling*: antes da ativação a leitura do FUXA acompanha o Gêmeo Digital; após a interceptação, o valor exibido congela em 50% enquanto o nível real atinge 100%: uma divergência de 50 pontos percentuais sem qualquer alarme nativo do SCADA.

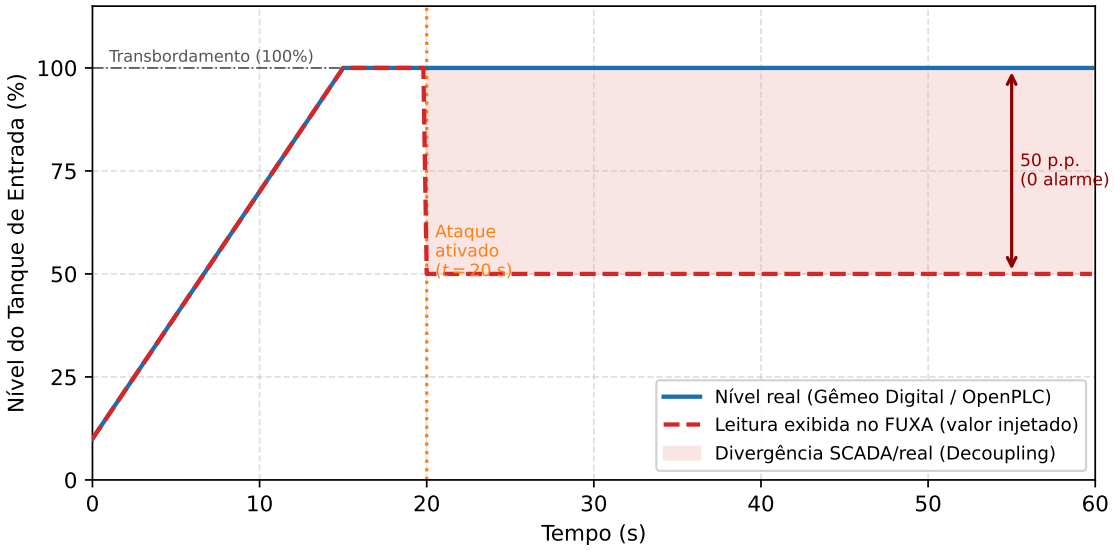

Figura 3. Evolução temporal do nível do Tanque de Entrada: valor real (Gêmeo Digital/OpenPLC) versus leitura injetada ao FUXA.

4.2. Eficácia da Defesa Ativa

O `detect_mitm.py` (T1) disparou `[!!!] ALERTA DE SEGURANÇA OT [!!!]` imediatamente após a ativação dos Terminais 3 e 4, identificando os MACs forjados (Listagem 2).

Listing 2. Saída do `detect_mitm.py`.

```
[!!!] ALERTA DE SEGURANCA OT [!!!]
Possivel ARP Spoofing detectado no IP: 172.17.0.3
MAC Original: 26:98:43:7d:46:50  MAC Falso: 1a:2f:8f:30:d4:40
```

O tempo de neutralização de ≤ 500 ms decorre do período configurado para os anúncios GARP no script `defesa_ativa.py`, um limite superior teórico; a latência real de correção depende da implementação do *stack* ARP no sistema operacional do contêiner e não foi medida independentemente neste trabalho. Os pacotes GARP somam aproximadamente 1 kbps de sobrecarga de banda, valor calculado a partir do período de anúncio de 500 ms e do tamanho do quadro ARP (28 bytes), sem considerar cabeçalho Ethernet e preenchimento até o quadro mínimo de 64 bytes; o overhead real na rede é maior que esse cálculo. Nenhum dos dois valores considera *jitter* ou tráfego concorrente de uma rede OT real, com impacto esperado desprezível na estabilidade das transações Modbus TCP. A quantificação precisa desse impacto sob condições de rede física constitui avaliação pendente.

4.3. Discussão

Do ponto de vista operacional, o *Decoupling* é a ameaça mais relevante para concessionárias de saneamento: o operador recebe indicação de normalidade enquanto o tanque transborda (condição análoga à manipulação do NaOH em Oldsmar [1]) sem qualquer alarme nativo do SCADA. A imunização por GARP bidirecional mostrou-se eficaz mesmo sob descarte de pacotes por `iptables`, sem modificações nos CLPs de campo. Cabe notar que a abordagem é suscetível a *race conditions* caso o atacante aumente a

frequência de envenenamento; em tais cenários o intervalo GARP deve ser reduzido ou complementado por *Dynamic ARP Inspection* (DAI) em switches gerenciáveis.

Em redes com maior densidade de dispositivos e tráfego concorrente, o volume agregado de anúncios GARP tende a crescer proporcionalmente ao número de pares IP/-MAC protegidos, já que a defesa opera por difusão periódica e não por resposta sob demanda. Nas condições testadas, com um único par protegido, a sobrecarga foi marginal (aprox. 1 kbps); em segmentos industriais com dezenas de dispositivos e tráfego Modbus concorrente, esse custo escala linearmente e pode competir por banda com o tráfego de controle, o que motiva investigar limites de escalabilidade e estratégias de anúncio segmentado (por VLAN ou por grupo de ativos) como extensão futura da abordagem.

Do ponto de vista das limitações, o escopo experimental restringe-se a um único segmento de rede virtualizado com o atacante executando `arpspoof` padrão. Não foram avaliados: (i) cenários em que o atacante também emite GARPs a frequência superior (*race condition*); (ii) *replay attacks* e *IP spoofing*; (iii) protocolos além do Modbus TCP (DNP3, EtherNet/IP); (iv) variação de latência e *jitter* de redes OT físicas. As métricas de desempenho da defesa constituem estimativas calculadas a partir de parâmetros de configuração, não medições empíricas independentes.

5. Conclusão

Este trabalho demonstrou experimentalmente as três contribuições declaradas. A arquitetura *open-source* (OpenPLC no *host* físico e contêineres FUXA e Kali Linux em Docker) reproduz com fidelidade suficiente o comportamento de uma ETE para fins de análise de segurança do protocolo Modbus. Ataques MitM com manipulação de atuador e FDI via Modbus TCP foram executados com êxito, induzindo *Decoupling* total sem detecção nativa pelo SCADA. A Defesa Ativa por GARP bidirecional neutralizou ambos os vetores de ataque nas condições testadas, com latência de detecção inferior a 500 ms, valor que constitui limite inferior, uma vez que os experimentos foram conduzidos em rede virtualizada sem *jitter* industrial. Essa propriedade é relevante para redes OT com CLPs que não suportam TLS, embora a robustez da contramedida frente a atacantes adaptativos (e.g., *race conditions* por GARP a frequência superior) permaneça como questão em aberto. O caráter reproduzível e de custo próximo de zero torna o *cyber range* acessível a equipes sem financiamento industrial. Scripts, imagens Docker e lógica ST estão disponibilizados publicamente em repositório para consulta (Seção 3). Como trabalhos futuros, planejam-se a expansão do Gêmeo Digital para os seis estágios completos de uma ETE, IDS por invariantes de processo [4] e a reprodução em planta física, com medição direta do tempo de neutralização e da sobrecarga de banda do GARP sob *jitter* real.

Agradecimentos

Os autores agradecem à Universidade Federal de Itajubá (UNIFEI) pelo apoio institucional a este trabalho e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de Iniciação Tecnológica (PIBIT, processo 138524/2025-4) concedida no ciclo 2025–2026. Agradecem também à Clavis Segurança da Informação pelo apoio concedido.

Referências

[1] CISA (2021). Compromise of U.S. water treatment facility. ICS-CERT Alert AA21-042A. https://www.cisa.gov/uscert/ncas/alerts/aa21-042a.

[2] Beresford, D. (2011). Exploiting Siemens Simatic S7 PLCs. *Black Hat USA 2011*. https://media.blackhat.com/bh-us-11/Beresford/BH_US11_Beresford_S7_PLCs_WP.pdf.

[3] Liu, Y., Ning, P., and Reiter, M. K. (2011). False data injection attacks against state estimation in electric power grids. *ACM Trans. Inf. Syst. Secur.*, 14(1). DOI: https://doi.org/10.1145/1952982.1952995.

[4] Mathur, A. P. and Tippenhauer, N. O. (2016). SWaT: A water treatment testbed for research and training on ICS security. Em *Proc. CySWater 2016*, pp. 31–36. IEEE. DOI: https://doi.org/10.1109/CySWater.2016.7469060.

[5] Morris, T. and Gao, W. (2014). Industrial control system traffic data sets for intrusion detection research. Em *Critical Infrastructure Protection VIII*, IFIP AICT vol. 441, pp. 65–78. Springer. DOI: https://doi.org/10.1007/978-3-662-45355-1_5.

[6] Stouffer, K. et al. (2023). Guide to operational technology (OT) security. NIST SP 800-82 Rev. 3. DOI: https://doi.org/10.6028/NIST.SP.800-82r3.

[7] Alves, F. et al. (2022). Exploiting vulnerabilities in the SCADA Modbus protocol. Em *Lecture Notes in Computer Science*, vol. 13204. Springer.

[8] Sanchez, G. (2020). Man-in-the-middle attack against Modbus TCP. Relatório técnico, Semantic Scholar. Acesso: 09 mai. 2026.

[9] Fritz, R. T. (2017). Melhoria da automação dos processos de uma ETE. Projeto de Graduação, UFRJ.

[10] Bruschi, D., Ornaghi, A., and Rosti, E. (2003). S-ARP: A secure address resolution protocol. Em *Proc. ACSAC 2003*, pp. 66–74. IEEE. DOI: https://doi.org/10.1109/CSAC.2003.1254311.