

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Engenharia Elétrica
Programa de Pós-Graduação em Engenharia Elétrica

Guilherme Ferreira César

**VULNERABILIDADES TEMPORAIS EM SISTEMAS DE
CONTROLE DIGITAL: O ATAQUE DE SUPERAMOSTRAGEM
E SUA DETECÇÃO**

Belo Horizonte
2026

Guilherme Ferreira César

**VULNERABILIDADES TEMPORAIS EM SISTEMAS DE CONTROLE
DIGITAL: O ATAQUE DE SUPERAMOSTRAGEM E SUA DETECÇÃO**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Engenharia Elétrica.

Área de concentração: Sinais e Sistemas

Linha de pesquisa: Controle e Modelagem de Sistemas Ciberfísicos

Orientador: Víctor Costa da Silva Campos

Coorientador: Luis Antonio Aguirre

Belo Horizonte

2026

C421v	César, Guilherme Ferreira. Vulnerabilidades temporais em sistemas de controle digital [recurso eletrônico] : o ataque de superamostragem e sua detecção / Guilherme Ferreira César. – 2026. 1 recurso online (155 f. : il., color.) : pdf. Orientador: Víctor Costa da Silva Campos. Coorientador: Luís Antônio Aguirre. Dissertação (mestrado) – Universidade Federal de Minas Gerais, Escola de Engenharia. Inclui bibliografia. 1. Engenharia elétrica – Teses. 2. Sistemas lineares – Teses. 3. Cibernética – Teses. 4. Detecção de anomalias (Computação) – Teses. I. Campos, Víctor Costa da Silva. II. Aguirre, Luís Antônio. III. Universidade Federal de Minas Gerais. Escola de Engenharia. IV. Título. CDU: 621.3(043)
-------	---



UNIVERSIDADE FEDERAL DE MINAS GERAIS
Escola de Engenharia
COLEGIADO DO CURSO DE PÓS-GRADUAÇÃO EM Engenharia Elétrica

FOLHA DE APROVAÇÃO

"Vulnerabilidades Temporais em Sistemas de Controle Digital: O Ataque de Superamostragem e Sua Detecção"

Guilherme Ferreira César

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Mestre em Engenharia Elétrica.

Aprovada em 05 de fevereiro de 2026.

Por:

Prof. Dr. Víctor Costa da Silva Campos
DELT (UFMG) - Orientador

Prof. Dr. Fernando de Oliveira Souza
DELT (UFMG)

Prof. Dr. Leandro Freitas de Abreu
DELT (UFMG)



Documento assinado eletronicamente por **Victor Costa da Silva Campos, Professor do Magistério Superior**, em 06/02/2026, às 16:50, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Leandro Freitas de Abreu, Professor do Magistério Superior**, em 06/02/2026, às 16:52, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Fernando de Oliveira Souza, Professor do Magistério Superior**, em 06/02/2026, às 18:06, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4915651** e o código CRC **C529D59B**.

Referência: Processo nº 23072.204816/2026-18

SEI nº 4915651

A Deus.

Agradecimentos

“Tua é, Senhor, a magnificência, e o poder, e a honra, e a vitória, e a majestade; porque teu é tudo quanto há nos céus e na terra; teu é, Senhor, o reino, e tu te exaltaste sobre todos como chefe. E riquezas e glória vêm de diante de ti, e tu dominas sobre tudo, e na tua mão há força e poder; e na tua mão está o engrandecer e dar força a tudo” (1 Cr 29:11-12)

A Deus, primeiramente, toda honra e toda glória. Nada do que está registrado nestas páginas seria possível sem a Sua graça e provisão. Foi Ele quem sustentou cada passo desta caminhada, nos momentos de alegria e, especialmente, nos momentos de dificuldade. Encerro este mestrado com a certeza de que não cheguei até aqui por mérito próprio, mas porque Ele, em Sua fidelidade, não me abandonou.

À minha família, expresso minha gratidão pelo apoio incondicional ao longo deste período. O suporte emocional e financeiro que me ofereceram foi alicerce indispensável para que eu pudesse me dedicar a este trabalho. Cada gesto de cuidado e cada palavra de encorajamento fizeram diferença real nesta jornada.

À minha companheira, Kamylla Luísa, agradeço de coração pelo apoio constante, pelas horas de sono perdidas ao meu lado, por me ouvir nos momentos de angústia e por oferecer conselho preciso nas horas certas. Deus a colocou em meu caminho, e sou grato a Ele por essa dádiva. Sua presença tornou este percurso mais leve e mais humano.

Ao meu orientador, Prof. Víctor Costa da Silva Campos, e ao meu coorientador, Prof. Luis Antonio Aguirre, registro meu sincero agradecimento pelos ensinamentos, pela paciência, pelas correções precisas e pelos conselhos que moldaram não apenas este trabalho, mas minha formação como pesquisador. A dedicação de ambos foi fundamental para o resultado aqui apresentado.

À Universidade Federal de Minas Gerais, pela oportunidade de formação em uma instituição de excelência, e ao Programa de Pós-Graduação em Engenharia Elétrica (PPGEE), pela estrutura, pelo suporte acadêmico e pela confiança depositada em meu trabalho ao longo deste período.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro que viabilizou a realização desta pesquisa.

*“Ora, ao Rei dos séculos, imortal, invisível, ao único Deus seja honra e glória para todo o
sempre. Amém.”*

(Bíblia Sagrada, 1 Timóteo 1:17, ARC)

RESUMO

Sistemas ciberfísicos combinam processos físicos com controle computacional, sendo fundamentais para infraestruturas críticas. A dependência crescente de redes digitais amplia a exposição a sofisticados ataques cibernéticos, cujas consequências podem incluir danos físicos e interrupção de serviços essenciais. Esta dissertação investiga vulnerabilidades inerentes ao comportamento dinâmico nos intervalos entre amostras de sistemas de controle digital. Apresenta-se o *ataque de superamostragem*, que manipula os atuadores da planta física entre os instantes de amostragem, mantendo as medições discretas aparentemente normais nos instantes de amostragem. Essa característica torna o ataque invisível a detectores tradicionais baseados em resíduos. O trabalho desenvolve a formulação matemática completa do ataque para sistemas lineares e não lineares, modelado como problema de otimização sob restrições de furtividade e maximização de impacto energético. Para sistemas não lineares, duas abordagens complementares são apresentadas: a modelagem *fuzzy Takagi-Sugeno*, que preserva a interpretabilidade física por meio da interpolação de subsistemas lineares locais, e a teoria dos Operadores de *Koopman*, que representa o sistema em um espaço de observáveis de alta dimensão onde a dinâmica se torna globalmente linear. Como estratégia de defesa, propõe-se um método de detecção baseado em estimação de estados operando com frequência superior à do controlador, utilizando detector estatístico χ^2 e análise de autocorrelação. A eficácia da abordagem é validada experimentalmente em três cenários de complexidade crescente. O Cenário 1 valida computacionalmente o ataque em um sistema linear SISO de segunda ordem (integrador duplo) e em um sistema linear MIMO de quarta ordem (modelo longitudinal da aeronave F-8C DFBW), demonstrando sua capacidade de permanecer indetectável aos métodos convencionais enquanto aumenta significativamente o consumo energético do sistema. O Cenário 2 estende a formulação para sistemas não lineares, utilizando um motor de corrente contínua com saturação magnética como caso de estudo, onde as abordagens *Takagi-Sugeno* e *Koopman* são implementadas. O Cenário 3 apresenta a implementação física na planta piloto LAB-Volt 3503-M0, adotando a perspectiva do atacante desde o reconhecimento do sistema até a execução do ataque, incluindo identificação de modelos, análise de vulnerabilidades e experimentos com diferentes tipos de ataques ciberfísicos (injeção de dados falsos, *replay*, *delay* e superamostragem). Os resultados confirmam tanto a capacidade destrutiva do ataque de superamostragem quanto a confiabilidade do método de detecção proposto.

Palavras-chave: sistemas ciberfísicos; ataques furtivos; segurança temporal; filtro de *Kalman*; infraestruturas críticas; controle digital; *Takagi-Sugeno*; operadores de *Koopman*.

ABSTRACT

Cyber-physical systems combine physical processes with computational control and are fundamental to critical infrastructure. The growing dependence on digital networks increases exposure to sophisticated cyberattacks, whose consequences may include physical damage and disruption of essential services. This dissertation investigates vulnerabilities inherent to dynamic behavior in the intervals between samples of digital control systems. The *supersampling attack* is presented, which manipulates the actuators of the physical plant between sampling instants, keeping discrete measurements apparently normal at sampling instants. This characteristic makes the attack invisible to traditional residual-based detectors. The work develops the complete mathematical formulation of the attack for linear and nonlinear systems, modeled as an optimization problem under stealth constraints and energy impact maximization. For nonlinear systems, two complementary approaches are presented: Takagi-Sugeno fuzzy modeling, which preserves physical interpretability through interpolation of local linear subsystems, and Koopman Operator theory, which represents the system in a high-dimensional observable space where dynamics become globally linear. As a defense strategy, a detection method based on state estimation operating at a higher frequency than the controller is proposed, utilizing a χ^2 statistical detector and autocorrelation analysis. The effectiveness of the approach is experimentally validated in three scenarios of increasing complexity. Scenario 1 computationally validates the attack on a second-order SISO linear system (double integrator) and on a fourth-order MIMO linear system (longitudinal model of F-8C DFBW aircraft), demonstrating its ability to remain undetectable to conventional methods while significantly increasing system energy consumption. Scenario 2 extends the formulation to nonlinear systems, using a direct current motor with magnetic saturation as a case study, where Takagi-Sugeno and Koopman approaches are implemented. Scenario 3 presents the physical implementation on the LAB-Volt 3503-M0 pilot plant, adopting the attacker's perspective from system reconnaissance to attack execution, including model identification, vulnerability analysis, and experiments with different types of cyber-physical attacks (false data injection, replay, delay, and supersampling). The results confirm both the destructive capability of the supersampling attack and the reliability of the proposed detection method.

Keywords: cyber-physical systems; stealthy attacks; temporal security; Kalman filter; critical infrastructures; digital control; Takagi-Sugeno; Koopman operators.

Lista de ilustrações

Figura 1 – Esquema geral da organização dos capítulos da dissertação.	27
Figura 2 – Arquitetura de sistema de controle sob ataque FDI.	29
Figura 3 – Arquitetura de sistema de controle sob ataque de repetição.	34
Figura 4 – Arquitetura de sistema de controle sob ataque de atraso temporal. . . .	38
Figura 5 – Arquitetura de sistema de controle sob ataque de espionagem.	43
Figura 6 – Diagrama de blocos do sistema de controle sob ataque de superamostragem.	47
Figura 7 – Evolução temporal da variável de estado x_1 para o problema de regulação com presença de ruído.	72
Figura 8 – Evolução temporal da variável de estado x_2 para o problema de regulação com presença de ruído.	72
Figura 9 – <i>Zoom</i> da evolução temporal da variável de estado x_1 para o problema de regulação com presença de ruído.	73
Figura 10 – <i>Zoom</i> da evolução temporal da variável de estado x_2 para o problema de regulação com presença de ruído.	73
Figura 11 – Ação de controle para o problema de regulação com presença de ruído.	74
Figura 12 – Evolução temporal da variável de estado x_1 para o problema de rastreamento com presença de ruído.	76
Figura 13 – Evolução temporal da variável de estado x_2 para o problema de rastreamento com presença de ruído.	76
Figura 14 – <i>Zoom</i> da evolução temporal da variável de estado x_1 para o problema de rastreamento com presença de ruído.	77
Figura 15 – <i>Zoom</i> da evolução temporal da variável de estado x_2 para o problema de rastreamento com presença de ruído.	77
Figura 16 – Ação de controle para o problema de rastreamento com presença de ruído.	78
Figura 17 – Resultados da detecção do ataque de superamostragem para o problema de regulação.	81
Figura 18 – Resultados da detecção do ataque de superamostragem para o problema de rastreamento.	83
Figura 19 – Evolução temporal das variáveis de estado da aeronave F-8C DFBW. . .	86
Figura 20 – Deflexões das superfícies de controle da aeronave F-8C DFBW.	86
Figura 21 – Evolução temporal das variáveis de estado do motor DC.	93
Figura 22 – Ação de controle aplicada ao motor DC.	93
Figura 23 – Variáveis de estado do motor DC sob ataque via <i>Koopman</i> (biblioteca “pobre”).	98
Figura 24 – Variáveis de estado do motor DC sob ataque via <i>Koopman</i> (biblioteca “rica”).	99

Figura 25 – Ação de controle do ataque via <i>Koopman</i> (biblioteca “pobre”).	99
Figura 26 – Ação de controle do ataque via <i>Koopman</i> (biblioteca “rica”).	100
Figura 27 – Erro de coincidência do ataque via <i>Koopman</i> (biblioteca “pobre”).	100
Figura 28 – Erro de coincidência do ataque via <i>Koopman</i> (biblioteca “rica”).	101
Figura 29 – Diagrama P&ID da planta piloto LAB-Volt 3503-M0.	102
Figura 30 – Topologia normal de comunicação.	105
Figura 31 – Topologia de comunicação após ataque <i>man-in-the-middle</i>	105
Figura 32 – Dados coletados durante experimento de identificação.	107
Figura 33 – Comparação entre a saída medida e as saídas dos modelos identificados.	109
Figura 34 – Análise de resíduos e autocorrelação dos modelos identificados.	110
Figura 35 – Validação do modelo ARMAX selecionado.	111
Figura 36 – Evolução temporal da saída para o problema de rastreamento com presença de ruído.	115
Figura 37 – Ação de controle para o problema de rastreamento com presença de ruído.	115
Figura 38 – Evolução temporal da saída da válvula nos experimentos físicos.	118
Figura 39 – Ação de controle aplicada à válvula nos experimentos físicos.	119
Figura 40 – Resultados da detecção do ataque de superamostragem nos experimentos físicos utilizando a função de autocorrelação.	120
Figura 41 – Resultados da detecção do ataque de superamostragem nos experimentos físicos utilizando a estatística qui-quadrado.	121
Figura 42 – Resposta do sistema sob ataque FDI aditivo.	143
Figura 43 – Resposta do sistema sob ataque FDI multiplicativo.	144
Figura 44 – Resposta do sistema sob ataque FDI senoidal.	145
Figura 45 – Espectrograma do sinal de saída durante operação normal.	147
Figura 46 – Análise do sinal para seleção da janela de gravação.	148
Figura 47 – Efeito do ataque de <i>replay</i> na saída do sistema	150
Figura 48 – Ação de controle durante o ataque de <i>replay</i>	151
Figura 49 – Análise de estabilidade em função do atraso para diferentes ganhos.	153
Figura 50 – Migração dos autovalores com o aumento do atraso.	154
Figura 51 – Resposta do sistema sob ataque de atraso.	155

Lista de tabelas

Tabela 1 – Comparação entre formulações <i>Takagi-Sugeno</i> e Operadores de <i>Koopman</i> para ataque de superamostragem.	67
Tabela 2 – Energia consumida e índice de desgaste do atuador para diferentes fatores de superamostragem no problema de regulação com ruído ($\sigma = 0,2$).	79
Tabela 3 – Energia consumida e índice de desgaste do atuador para diferentes fatores de superamostragem no problema de rastreamento com ruído ($\sigma = 0,2$).	79
Tabela 4 – Parâmetros da simulação do ataque MIMO no sistema F-8C DFBW.	85
Tabela 5 – Métricas de impacto do ataque de superamostragem no sistema F-8C DFBW.	87
Tabela 6 – Parâmetros do motor DC com saturação magnética.	88
Tabela 7 – Configuração da simulação do ataque no motor DC.	89
Tabela 8 – Ganhos do controlador DLQR para cada regra T-S.	91
Tabela 9 – Métricas de impacto do ataque de superamostragem no motor DC via modelo T-S.	92
Tabela 10 – Configuração da identificação EDMD para o motor DC.	94
Tabela 11 – Comparação das métricas de impacto do ataque via <i>Koopman</i> para diferentes bibliotecas.	97
Tabela 12 – Comparação dos modelos candidatos para identificação do sistema.	108
Tabela 13 – Parâmetros da simulação do ataque de superamostragem.	114
Tabela 14 – Métricas de impacto do ataque de superamostragem.	116
Tabela 15 – Parâmetros dos experimentos de validação física do ataque de superamostragem.	117
Tabela 16 – Métricas de impacto do ataque de superamostragem nos experimentos físicos.	119
Tabela 17 – Métricas de desempenho sob diferentes ataques FDI.	146
Tabela 18 – Influência do ganho do controlador na robustez ao ataque de atraso.	152
Tabela 19 – Análise de estabilidade do sistema com atraso ($R = 10$).	153

Lista de abreviaturas e siglas

ACF	<i>Autocorrelation Function</i>
ARP	<i>Address Resolution Protocol</i>
BDD	<i>Bad Data Detection</i>
CPS	<i>Cyber-Physical Systems</i>
DC	<i>Direct Current</i>
DCS	<i>Distributed Control Systems</i>
DFBW	<i>Digital Fly-By-Wire</i>
DFT	<i>Discrete Fourier Transform</i>
DLQR	<i>Discrete Linear Quadratic Regulator</i>
DoS	<i>Denial of Service</i>
EDMD	<i>Extended Dynamic Mode Decomposition</i>
ELM	<i>Extreme Learning Machine</i>
FDI	<i>False Data Injection</i>
FFT	<i>Fast Fourier Transform</i>
GPS	<i>Global Positioning System</i>
GRASP	<i>Greedy Randomized Adaptive Search Procedure</i>
IIoT	<i>Industrial Internet of Things</i>
IoT	<i>Internet of Things</i>
IP	<i>Internet Protocol</i>
LPV	<i>Linear Parameter-Varying</i>
LQR	<i>Linear Quadratic Regulator</i>
LSTM	<i>Long Short-Term Memory</i>
LTI	<i>Linear Time-Invariant</i>

MAC	<i>Media Access Control</i>
MIMO	<i>Multiple-Input Multiple-Output</i>
NASA	<i>National Aeronautics and Space Administration</i>
P&ID	<i>Piping and Instrumentation Diagram</i>
PCA	<i>Principal Component Analysis</i>
PDC	<i>Parallel Distributed Compensation</i>
PLC	<i>Programmable Logic Controllers</i>
PRBS	<i>Pseudo-Random Binary Sequence</i>
RBF	<i>Radial Basis Function</i>
RCL	<i>Restricted Candidate List</i>
RMS	<i>Root Mean Square</i>
RTU	<i>Remote Terminal Unit</i>
SCADA	<i>Supervisory Control and Data Acquisition</i>
SDP	<i>Semidefinite Programming</i>
SISO	<i>Single-Input Single-Output</i>
SPRT	<i>Sequential Probability Ratio Test</i>
SQL	<i>Structured Query Language</i>
STFT	<i>Short-Time Fourier Transform</i>
SVD	<i>Singular Value Decomposition</i>
T-S	<i>Takagi-Sugeno</i>
TCO	<i>Total Cost of Ownership</i>
TCP	<i>Transmission Control Protocol</i>
UART	<i>Universal Asynchronous Receiver-Transmitter</i>
USB	<i>Universal Serial Bus</i>
VPN	<i>Virtual Private Network</i>
WLS	<i>Weighted Least Squares</i>
ZOH	<i>Zero-Order Hold</i>

Sumário

1	INTRODUÇÃO	17
1.1	Tipos de Ataques	19
1.2	Justificativa e Relevância	23
1.3	Objetivos Gerais e Específicos	24
1.4	Publicações	25
1.5	Organização do Texto	25
2	FUNDAMENTAÇÃO TEÓRICA	28
2.1	Ataques de Injeção de Dados Falsos	28
2.2	Ataques de Repetição	33
2.3	Ataques de Atraso Temporal	37
2.4	Ataques de Espionagem	42
3	FORMULAÇÃO DO ATAQUE DE SUPERAMOSTRAGEM E METODOLOGIA DE DETECÇÃO	47
3.1	Definição do Problema	47
3.2	Formulação do Ataque em Sistemas Lineares	48
3.3	Extensão para Sistemas Não Lineares	57
4	RESULTADOS E PROCEDIMENTOS EXPERIMENTAIS	70
4.1	Cenário 1: Validando o Ataque de Superamostragem Computacionalmente em Sistemas Lineares	70
4.2	Cenário 2: Explorando o Ataque de Superamostragem em Sistemas Não Lineares	87
4.3	Cenário 3: Perspectiva do Atacante - Reconhecimento, Preparação e Execução do Ataque	98
5	CONCLUSÕES	122
5.1	Síntese do Trabalho	122
5.2	Principais Contribuições	123
5.3	Limitações Identificadas	123
5.4	Perspectivas para Pesquisas Futuras	125
5.5	Considerações Finais	128
	REFERÊNCIAS	129

APÊNDICES **137**

APÊNDICE A – ALGORITMOS IMPLEMENTADOS 138

- A.1 Ataque de Superamostragem T-S MIMO via GRASP (C-GRASP) . . . 138
- A.2 Ataque de Superamostragem MIMO via *Koopman* (EDMD-GRASP) . . 139

APÊNDICE B – IMPLEMENTAÇÃO EXPERIMENTAL DE ATAQUES CIBERNÉTICOS CLÁSSICOS 142

- B.1 Ataque de Injeção de Dados Falsos 142
- B.2 Ataque de *Replay* 146
- B.3 Ataque de Atraso 151

1 Introdução

A era dos sistemas ciberfísicos (CPS, do inglês *cyber-physical systems*) marca uma mudança significativa na forma como o mundo físico e digital se relacionam. Esses sistemas nascem da combinação entre computação, redes de comunicação e elementos físicos, possibilitando que dispositivos conectados capturem informações, as processem e ajam sobre elas rapidamente. Essa integração quebrou as barreiras que antes separavam a computação do ambiente físico, criando a base para tecnologias inteligentes e conectadas, que hoje estão presentes desde fábricas e indústrias até infraestruturas essenciais como energia, transporte, saneamento e defesa (LEE, 2008; BAHETI; GILL, 2011).

O conceito de CPS se concretiza na Indústria 4.0, um modelo industrial baseado em automação avançada, sensores distribuídos, comunicação entre máquinas e análise de grandes volumes de dados. Como apontam Sanislav e Miclea (2012), esses sistemas funcionam em ciclos contínuos de monitoramento, decisão e ação, integrando de forma adaptativa as camadas físicas e virtuais. Porém, essa integração, apesar de trazer ganhos significativos de eficiência e produtividade, também traz novos riscos relacionados à crescente dependência de redes de comunicação, *software* e da necessidade de diferentes plataformas trabalharem juntas.

A complexidade das arquiteturas de controle industrial modernas e a conexão entre sistemas que antes funcionavam isolados aumentaram consideravelmente a exposição dos ambientes produtivos a ataques. Tradicionalmente, redes industriais baseadas em sistemas de supervisão e aquisição de dados (SCADA, do inglês *supervisory control and data acquisition*), sistema digital de controle distribuído (DCS, do inglês *distributed control system*) e controlador lógico programável (PLC, do inglês *programmable logic controller*) operavam em ambientes separados, usando protocolos proprietários e comunicação local. Porém, a necessidade de integração com redes corporativas e com a *Internet*, motivada pela demanda por monitoramento remoto, manutenção preditiva e gestão centralizada, acabou com o isolamento que oferecia certa proteção a esses sistemas. Como mostram Rubio et al. (2019), essa conectividade tornou os sistemas industriais vulneráveis a ameaças cibernéticas que antes se limitavam ao campo da tecnologia da informação.

Como apontam Nankya, Chataut e Akl (2023), a conexão de dispositivos inteligentes por meio do conceito de *Internet* industrial das coisas (IIoT, do inglês *industrial Internet of things*) criou uma arquitetura em múltiplas camadas formada por sensores, atuadores, *gateways* e plataformas em nuvem, cujas comunicações dependem de protocolos abertos que, muitas vezes, não foram projetados para resistir a ataques sofisticados. O resultado é um ambiente onde a vulnerabilidade de um único componente pode comprometer toda

a cadeia de produção. Esse problema se agrava pela presença de equipamentos antigos e pela falta de mecanismos robustos de autenticação e criptografia (ALNAJIM et al., 2023).

O aumento dessas interdependências exige repensar a segurança cibernética de forma mais ampla. Não se trata apenas de proteger dados, mas de garantir a integridade e a disponibilidade de processos físicos que dependem de informação digital. Um ataque cibernético a um CPS pode ter consequências reais: danos a equipamentos, interrupção de serviços essenciais ou até riscos à vida humana. Estudos recentes mostram que, ao contrário dos ambientes puramente digitais, nos CPS o impacto das falhas de segurança vai além do ciberespaço (domínio global formado por todas as infraestruturas interconectadas de tecnologia da informação, redes de telecomunicações, sistemas computacionais e controladores embutidos (ZDZIKOT, 2021)) e se manifesta em eventos concretos, como explosões industriais, contaminação de água ou apagões generalizados (HARKAT et al., 2024; RUBIO et al., 2019).

Essa vulnerabilidade ficou clara com a popularização de dispositivos industriais baseados em redes de protocolo de *Internet* (IP, do inglês *Internet protocol*) e a crescente integração de tecnologias emergentes como computação em nuvem e aprendizado de máquina. Embora a automação inteligente traga benefícios significativos, ela também cria novos vetores de ataque, já que cada camada de conectividade adiciona dependências que podem ser exploradas por agentes maliciosos (NANKYA; CHATAUT; AKL, 2023).

Nesse cenário, a segurança dos CPS depende não apenas de soluções técnicas, mas de uma mudança cultural e organizacional. A defesa deve abranger desde políticas de controle de acesso e segmentação de redes até a implementação de monitoramento contínuo com base em técnicas de detecção de anomalias. Pesquisas recentes apontam para o uso de algoritmos de aprendizado de máquina como instrumentos promissores na identificação de comportamentos anômalos e na predição de falhas de segurança (NANKYA; CHATAUT; AKL, 2023).

No entanto, compreender a evolução das vulnerabilidades ciberfísicas exige também analisar eventos históricos que mudaram a percepção global sobre segurança industrial. Um dos casos mais marcantes é o incidente conhecido como *Maroochy Water Breach*, ocorrido em 2000, na Austrália. Nesse episódio, um ex-funcionário insatisfeito da empresa *Maroochy Shire Council* usou um computador portátil e um transmissor de rádio para acessar o sistema de controle de uma estação de tratamento de esgoto, liberando milhões de litros de efluente não tratado, o que causou graves danos ambientais e sanitários (ABRAMS; WEISS, 2008). Esse ataque mostrou o potencial destrutivo de ações direcionadas a sistemas de controle e marcou o início de uma nova compreensão sobre a necessidade de segurança cibernética em infraestruturas críticas.

Outro marco foi a descoberta do *worm Stuxnet*, identificado em 2010. Diferente de vírus comuns, o *Stuxnet* foi um código malicioso sofisticado, projetado para atacar

controladores industriais da *Siemens* usados na usina nuclear de Natanz, no Irã. Foi o primeiro *malware* conhecido a causar danos físicos diretos a equipamentos industriais, marcando o início da era das armas cibernéticas (LANGNER, 2011). O *Stuxnet* se infiltrava em computadores por meio de dispositivos de barramento serial universal (USB, do inglês *universal serial bus*) e, uma vez dentro da rede, modificava o comportamento dos controladores que regulavam a velocidade das centrífugas, enquanto mostrava dados falsos aos operadores (AANONSEN, 2025). O ataque mostrou que *softwares* maliciosos podiam ir além do ambiente digital e operar diretamente sobre processos físicos com precisão e propósitos geopolíticos.

Mais recentemente, os ataques contra a infraestrutura de energia da Ucrânia mostraram como ameaças ciberfísicas se concretizam em situações reais de conflito. Em 2015 e 2016, ataques atribuídos a grupos ligados à inteligência russa cortaram o fornecimento de energia elétrica para centenas de milhares de pessoas. Em 2022, um novo incidente relatado pela Mandiant revelou o uso de técnicas diferentes para comprometer sistemas de supervisão SCADA por meio de um *software* legítimo chamado *MicroSCADA*. O grupo responsável, conhecido como Sandworm, explorou falhas de um hipervisor que hospedava os controladores do sistema elétrico, inserindo comandos maliciosos que podiam abrir disjuntores e causar apagões coordenados (DRAGOS, 2023). O ataque usou o método “*Living off the Land*”, ou seja, usar ferramentas legítimas do próprio sistema para executar ações maliciosas, o que dificulta a detecção. Depois de atacar os sistemas de controle, os invasores lançaram o *malware CaddyWiper* para apagar rastros e destruir dados nos servidores de tecnologia (PROSKA et al., 2023).

Esses casos revelam um padrão comum: a ligação entre o mundo digital e o físico transforma a segurança cibernética em questão de segurança pública e soberania nacional. O *Maroochy Water Breach* mostrou que uma única pessoa poderia causar danos ambientais graves; o *Stuxnet* provou que países podem usar ataques cibernéticos como ferramentas de política internacional; e os ataques à infraestrutura ucraniana mostraram que operações ciberfísicas podem acontecer ao mesmo tempo que ataques militares convencionais (RUBIO et al., 2019; AANONSEN, 2025). A era dos CPS se caracteriza, então, pela combinação entre avanço tecnológico e exposição a novos riscos. O desafio atual é construir uma infraestrutura inteligente que seja, ao mesmo tempo, eficiente e segura, reconhecendo que cada linha de código e cada sensor físico fazem parte de um mesmo sistema de riscos e responsabilidades compartilhadas.

1.1 Tipos de Ataques

A segurança em CPS vai além da simples proteção de dados, abrangendo também a integridade, a disponibilidade e a confiabilidade dos processos físicos controlados digi-

talmente. Com a crescente integração entre o mundo físico e o cibernético, compreender como o atacante age e quais mecanismos o tornam bem-sucedido é essencial para o desenvolvimento de estratégias de defesa eficazes (PASQUALETTI; DÖRFLER; BULLO, 2013; MO; SINOPOLI, 2009; YANG et al., 2020).

De forma geral, um ataque ciberfísico segue três etapas principais: (1) reconhecimento e coleta de informações sobre a planta, (2) inserção ou manipulação de dados e (3) execução discreta para evitar a detecção. Essa última fase é o que diferencia os ataques modernos, baseados em modelagem, dos ataques tradicionais de negação de serviço (DoS, do inglês *denial of service*). Enquanto um ataque DoS interrompe o funcionamento do sistema de maneira evidente, os ataques furtivos buscam comprometer a integridade e a estabilidade sem chamar atenção, exigindo contramedidas mais avançadas.

O desafio da segurança nesses sistemas apresenta múltiplas dimensões que atravessam distintas camadas operacionais. Essas camadas abrangem desde a infraestrutura física até os protocolos de comunicação e sistemas de controle, todas potencialmente vulneráveis a adversários com diferentes níveis de sofisticação técnica. É fundamental compreender que o adversário não age de forma arbitrária ou oportunista. Ao contrário, trata-se de um operador estratégico que possui conhecimento sobre a dinâmica do sistema que pretende comprometer. Conforme demonstrado por Mo e Sinopoli (2016), Joo, Qu e Namerikawa (2021), a execução bem-sucedida de um ataque depende fundamentalmente de que o invasor tenha acesso a um modelo suficientemente preciso da planta industrial ou do sistema-alvo. Esse conhecimento pode ser obtido por diversos meios: por meio da interceptação de comunicações sensíveis, de processos de engenharia reversa, ou mesmo pela exploração de informações técnicas que foram disponibilizadas publicamente.

Na interceptação, o atacante observa passivamente o tráfego entre sensores, atuadores e controladores. A partir desses dados, é possível empregar técnicas de identificação de sistemas para reconstruir o comportamento dinâmico da planta (YANG et al., 2020). Esse tipo de espionagem é especialmente perigoso em sistemas de controle distribuído, nos quais múltiplos dispositivos comunicam-se em rede, pois a captura de uma pequena fração do tráfego pode ser suficiente para estimar as variáveis de estado mais relevantes. Em ataques de engenharia reversa, o adversário obtém acesso físico ou digital aos PLCs e extrai a lógica de controle ou parâmetros de *hardware*. Por outro lado, vulnerabilidades conhecidas em sistemas estabelecidos, especialmente em ambientes SCADA, podem ser exploradas sem grande esforço, dado que muitos ainda operam com protocolos antigos e sem autenticação.

Com esse conhecimento, o atacante torna-se capaz de projetar ações que não apenas interrompem o funcionamento do sistema, mas o fazem de maneira furtiva, escapando dos mecanismos tradicionais de detecção. Esses ataques discretos são denominados ataques furtivos e constituem um dos maiores desafios na segurança de CPS (MO; SINOPOLI, 2016;

GUO et al., 2022). Entre as técnicas mais estudadas encontram-se os ataques de injeção de dados falsos (FDI, do inglês *false data injection*), os ataques de repetição (*replay*), e os ataques temporais, que incluem atrasos deliberados (*delay attacks*) e a falsificação de sinais de tempo (*spoofing*).

No ataque de FDI, o invasor manipula dados de sensores ou atuadores, inserindo valores falsos que se mantêm coerentes com o modelo do sistema. Assim, o ataque passa despercebido pelos detectores clássicos de dados ruins (BDD, do inglês *bad data detection*), que verificam o resíduo entre as medições e as estimativas do estado (PASQUALETTI; DÖRFLER; BULLO, 2013; REN et al., 2023). Ao conhecer as equações que regem o sistema, o atacante ajusta os valores injetados de modo que o resíduo permaneça dentro dos limites normais, mesmo quando o sistema é levado a uma condição perigosa. Em termos simples, o sistema “acredita” estar operando normalmente, enquanto o processo físico caminha para a falha.

Os ataques de repetição, seguem uma lógica diferente. Nesse caso, o invasor grava medições reais em um período de operação normal e as retransmite posteriormente em laço fechado. Como os dados possuem estatísticas válidas, o controlador interpreta a situação como estável, enquanto o atacante manipula o atuador em paralelo (MO; SINOPOLI, 2009). Essa técnica não requer um conhecimento detalhado da planta, apenas a capacidade de capturar e reenviar pacotes, o que a torna uma ameaça real em sistemas com autenticação fraca ou sem criptografia.

Os ataques temporais exploram a dependência dos sistemas de controle em relação ao tempo e à sincronização. No ataque por atraso, os pacotes de dados são interceptados e reenviados com um atraso intencional, o que prejudica o desempenho do controlador e pode até levar o sistema à instabilidade (XIAHOU; LIU; WU, 2021). Em redes elétricas e em sistemas de grande escala que utilizam sinais do sistema de posicionamento global (GPS, do inglês *global positioning system*) para manter a sincronização, o ataque de *GPS spoofing* envia sinais falsos que parecem autênticos, provocando dessincronização entre as medições e as ações de controle, o que pode resultar em falhas generalizadas (LIANG et al., 2017).

Diante dessa variedade de ameaças, as estratégias de defesa evoluíram notadamente e podem ser divididas em duas categorias: abordagens passivas e ativas. As abordagens passivas têm como objetivo monitorar o comportamento do sistema por meio de observadores e detectores de anomalias. O método mais comum é o baseado em resíduos, que compara as saídas reais da planta com as saídas estimadas por um modelo matemático de referência, como um observador de estados ou um filtro de *Kalman*. Em condições normais, o resíduo se comporta como um ruído aleatório com média nula; quando ocorre um desvio relevante, um alarme é acionado (MEHRA; PESCHON, 1971; PASQUALETTI; DÖRFLER; BULLO, 2013). No entanto, como mostrado por Mo e Sinopoli (2016), ataques planejados com

conhecimento do modelo podem manter o resíduo dentro dos limites normais, tornando-se invisíveis para os detectores convencionais.

Como resposta a essa limitação, surgiram as abordagens ativas, nas quais o próprio sistema gera sinais de autenticação para verificar a integridade das medições. Um exemplo é o método de *watermarking*, em que o controlador insere um sinal aleatório de pequena amplitude no comando de controle. A planta responde a esse sinal e esse retorno é esperado nas medições; se o retorno não ocorrer, como acontece em um ataque de repetição, o sistema detecta a anomalia (MO; SINOPOLI, 2009; SATCHIDANANDAN; KUMAR, 2017). Esse método, embora efetivo, exige um cuidado para equilibrar a intensidade do sinal e a possível degradação do desempenho do controlador.

Outras técnicas ativas envolvem o uso de esquemas de codificação e mascaramento não lineares, desenvolvidos para confundir o atacante. De acordo com Joo, Qu e Namerikawa (2021), a inserção de componentes caóticas de codificação e decodificação nas comunicações pode impedir que o invasor compreenda a estrutura interna do sistema, tornando impossível reproduzir seu comportamento. De forma semelhante, Guo et al. (2022) propõem o uso de esquemas de codificação estocástica para proteger sistemas contra ataques furtivos, fazendo com que as variáveis trocadas entre a planta e o controlador sejam constantemente perturbadas por sinais aleatórios conhecidos apenas pelo sistema. Essa abordagem garante que, mesmo que o invasor intercepte o tráfego, ele não consiga deduzir o estado interno do sistema.

Em sistemas de energia elétrica, estratégias de compensação de atraso têm se mostrado promissoras contra ataques temporais. O método de controle robusto baseado em estimativa de perturbação utiliza um observador que estima o erro causado pelos atrasos e o compensa em tempo real (XIAHOU; LIU; WU, 2021). Essa abordagem tem a vantagem de manter o sistema estável mesmo quando ocorrem atrasos aleatórios introduzidos de forma maliciosa.

Esses avanços mostram que, embora os ataques estejam se tornando mais complexos, as defesas também evoluem com base no conhecimento do sistema e na modelagem matemática. Técnicas como observadores robustos, codificação dinâmica (técnica na qual os sinais transmitidos são continuamente transformados por esquemas variantes no tempo, impedindo que um invasor, mesmo interceptando o tráfego, consiga reconstruir o estado interno do sistema), controle resiliente e injeção de sinais autenticadores representam o que há de mais atual na proteção de sistemas ciberfísicos. O grande desafio está em combinar essas estratégias de forma que garantam segurança sem comprometer o desempenho, mantendo um equilíbrio entre robustez, custo computacional, custo operacional/energético e sensibilidade à detecção.

Compreender a segurança em sistemas ciberfísicos vai além de listar ataques e defesas, pois exige entender um ambiente em constante transformação, onde cada

avanço tecnológico pode gerar novas vulnerabilidades. A busca pela resiliência depende da integração entre controle, comunicação e segurança da informação, já que as fronteiras entre o mundo físico e o digital se tornam cada vez mais próximas.

1.2 Justificativa e Relevância

As estratégias de defesa apresentadas na seção anterior representam avanços importantes no campo da segurança dos CPS. No entanto, essas abordagens compartilham uma limitação conceitual importante: partem da premissa de que monitorar e proteger o sistema apenas no momento das medições digitais é suficiente para garantir sua segurança ao longo de toda a operação contínua. Embora essa hipótese seja conveniente do ponto de vista do projeto e da implementação digital, ela ignora a natureza híbrida dos CPS, que combinam variáveis contínuas e operações discretas. Essa limitação conceitual cria uma nova brecha de vulnerabilidade.

O trabalho de César, Campos e Aguirre (2025) formaliza e demonstra o aparecimento de um ataque que explora exatamente essa vulnerabilidade: o Ataque de Superamostragem (*Oversampling Attack*). Essa nova classe de ameaça quebra os princípios tradicionais em sistemas amostrados ao atuar entre as amostras, manipulando a dinâmica física da planta sem alterar as medições nos instantes de amostragem originais. Para isso, o ataque opera em uma taxa de amostragem superior à do controlador, o que lhe permite injetar sinais no atuador em instantes intermediários que o sistema de monitoramento convencional não observa. O sinal de ataque é projetado para atender a duas condições: a condição de furtividade, que assegura que as saídas da planta coincidam com as saídas esperadas nos instantes discretos, e a condição de impacto, que maximiza a energia entregue ao atuador durante o intervalo entre amostras.

A primeira condição garante que os mecanismos tradicionais de detecção, como observadores de estado, filtros de *Kalman* ou detectores de resíduos, não consigam identificar o ataque, uma vez que todas as medições nos instantes de amostragem parecem normais. A segunda revela a condição destrutiva do ataque, pois ao fazer o atuador oscilar em alta frequência, o invasor aumenta o consumo de energia e acelera o desgaste mecânico, o que pode causar falhas prematuras e elevar os custos de operação. Em termos práticos, o ataque não busca desestabilizar o sistema de controle, mas comprometer sua integridade física de forma silenciosa e contínua.

A importância desse tema fica clara ao observar que sistemas digitais de controle, usados em infraestruturas críticas e aplicações industriais, compartilham a mesma estrutura temporal explorada pelo ataque. Todo controlador digital opera com medições digitais, que definem quando as decisões de controle são tomadas. Entre essas medições, o sistema segue uma dinâmica contínua, criando uma oportunidade para que o atacante atue sem

ser detectado. Essa vulnerabilidade está ligada ao próprio conceito de controle digital, e não a uma implementação específica, o que amplia seu impacto.

Sob essa perspectiva, o estudo de César, Campos e Aguirre (2025) traz uma contribuição relevante ao revelar uma vulnerabilidade estrutural nos sistemas ciberfísicos e propor uma solução prática. O trabalho mostra que garantir segurança apenas nas medições digitais não é suficiente para proteger o sistema entre as medições. Como resposta, é apresentada uma estratégia de defesa baseada em um filtro de *Kalman* operando com uma taxa de amostragem mais alta que a do controlador, capaz de estimar as saídas do sistema nos instantes não medidos entre as amostras e, assim, detectar anomalias que passariam despercebidas. Os resultados experimentais demonstram que essa abordagem identifica o ataque de forma confiável, sem exigir mudanças no controlador principal.

A justificativa para esta dissertação está na necessidade de aprofundar o entendimento teórico e experimental sobre o ataque de superamostragem e suas implicações para a segurança de sistemas de controle amostrados. Ao apresentar um novo vetor de ataque, este trabalho amplia a análise de vulnerabilidades em sistemas ciberfísicos e propõe uma abordagem de defesa alinhada à natureza contínua do problema. A relevância da pesquisa está em mostrar que a segurança temporal, ou seja, a garantia de um comportamento seguro entre as amostras, deve ser tratada como um requisito de projeto tão importante quanto a estabilidade e o desempenho.

1.3 Objetivos Gerais e Específicos

Este trabalho tem como objetivo geral investigar vulnerabilidades temporais em sistemas de controle digital, com foco no Ataque de Superamostragem e em estratégias para sua detecção. A pesquisa busca demonstrar que a segurança de um sistema ciberfísico não pode ser avaliada apenas nos instantes de amostragem do controlador, sendo necessário considerar o comportamento do sistema nos intervalos entre amostras, onde ações maliciosas podem ocorrer sem deixar rastros nos mecanismos convencionais de monitoramento.

A dissertação desenvolve a formulação matemática do ataque para sistemas lineares e não lineares, propõe um método de detecção baseado em observador de estados operando em frequência superior à do controlador, e valida experimentalmente tanto a eficácia do ataque quanto a confiabilidade do detector.

Para alcançar esse propósito, os seguintes objetivos específicos são definidos:

- Apresentar os fundamentos teóricos da segurança em CPS, incluindo os principais tipos de ataque (FDI, *replay*, *delay* e espionagem) e as estratégias de defesa consolidadas na literatura, como detectores baseados em resíduos, *watermarking* e esquemas de codificação.

- Formalizar matematicamente o Ataque de Superamostragem para sistemas lineares, definindo o problema de otimização sob restrições de furtividade (coincidência das saídas nos instantes de amostragem) e maximização do impacto energético, além de desenvolver o mecanismo de detecção baseado em Filtro de *Kalman* superamostrado com detector estatístico χ^2 e análise da função de autocorrelação.
- Estender a formulação do ataque para sistemas não lineares por meio de duas abordagens complementares: a modelagem *fuzzy Takagi-Sugeno*, que preserva a interpretabilidade física por meio da interpolação de subsistemas lineares locais, e a teoria dos Operadores de *Koopman*, que eleva o sistema para um espaço de observáveis onde a dinâmica se torna globalmente linear.
- Validar computacionalmente o ataque e o método de detecção em sistemas lineares (integrador duplo e sistema MIMO de aeronave) e não lineares (motor de corrente contínua com saturação magnética).
- Implementar e executar experimentos físicos na planta piloto *LAB-Volt 3503-M0* (FESTO DIDACTIC, 2015), adotando a perspectiva completa do atacante desde o reconhecimento do sistema (identificação de modelos, análise de vulnerabilidades) até a execução de diferentes tipos de ataques ciberfísicos (FDI, *replay*, *delay* e superamostragem), validando a aplicabilidade prática das formulações desenvolvidas.

1.4 Publicações

Durante o desenvolvimento dessa dissertação, foi produzido e publicado o seguinte artigo, que apresenta parte dos resultados obtidos neste trabalho:

- César, Guilherme F.; Campos, Víctor C. da S.; Aguirre, Luis A. **Ataque de Superamostragem**. Artigo aceito para publicação em congresso. Trabalho apresentado no Simpósio Brasileiro de Automação Inteligente (SBAI 2025). Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais (UFMG), 2025.

1.5 Organização do Texto

O texto está organizado de forma a conduzir o leitor desde a contextualização teórica até a apresentação dos resultados experimentais e das conclusões do trabalho. A estrutura segue uma progressão lógica que parte dos fundamentos, passa pela formulação matemática do problema e culmina na validação experimental. A Figura 1 apresenta de forma esquemática essa organização.

O **Capítulo 1** apresenta a introdução do trabalho, contextualizando o crescimento dos sistemas ciberfísicos e as vulnerabilidades que surgem da integração entre o mundo físico e o digital. São discutidos os principais tipos de ataque e estratégias de defesa, a justificativa e relevância da pesquisa, os objetivos gerais e específicos, além das publicações relacionadas.

O **Capítulo 2** aborda a fundamentação teórica, detalhando os principais tipos de ataque em CPS: injeção de dados falsos (FDI), ataques de repetição (*replay*), ataques de atraso temporal (*delay*) e ataques de espionagem (*eavesdropping*). Para cada tipo de ataque, são apresentadas a modelagem matemática, as condições de furtividade e as estratégias de defesa correspondentes.

O **Capítulo 3** apresenta a formulação matemática do Ataque de Superamostragem e a metodologia de detecção. O capítulo desenvolve a definição formal do problema, a formulação do ataque em sistemas lineares, e a extensão para sistemas não lineares por meio das abordagens *Takagi-Sugeno* e Operadores de *Koopman*. Também é detalhado o método de detecção baseado em Filtro de *Kalman* superamostrado.

O **Capítulo 4** descreve os resultados e procedimentos experimentais, organizados em três cenários de complexidade crescente. O Cenário 1 valida computacionalmente o ataque em sistemas lineares (integrador duplo e sistema F-8C DFBW (ELLIOTT, 1977)). O Cenário 2 explora a extensão para sistemas não lineares utilizando um motor de corrente contínua com saturação magnética. O Cenário 3 apresenta a implementação física na planta piloto *LAB-Volt 3503-M0*, adotando a perspectiva do atacante desde o reconhecimento até a execução do ataque.

Por fim, o **Capítulo 5** apresenta as conclusões do trabalho, discutindo as principais contribuições, as limitações identificadas e as perspectivas para pesquisas futuras.

Organização da Dissertação

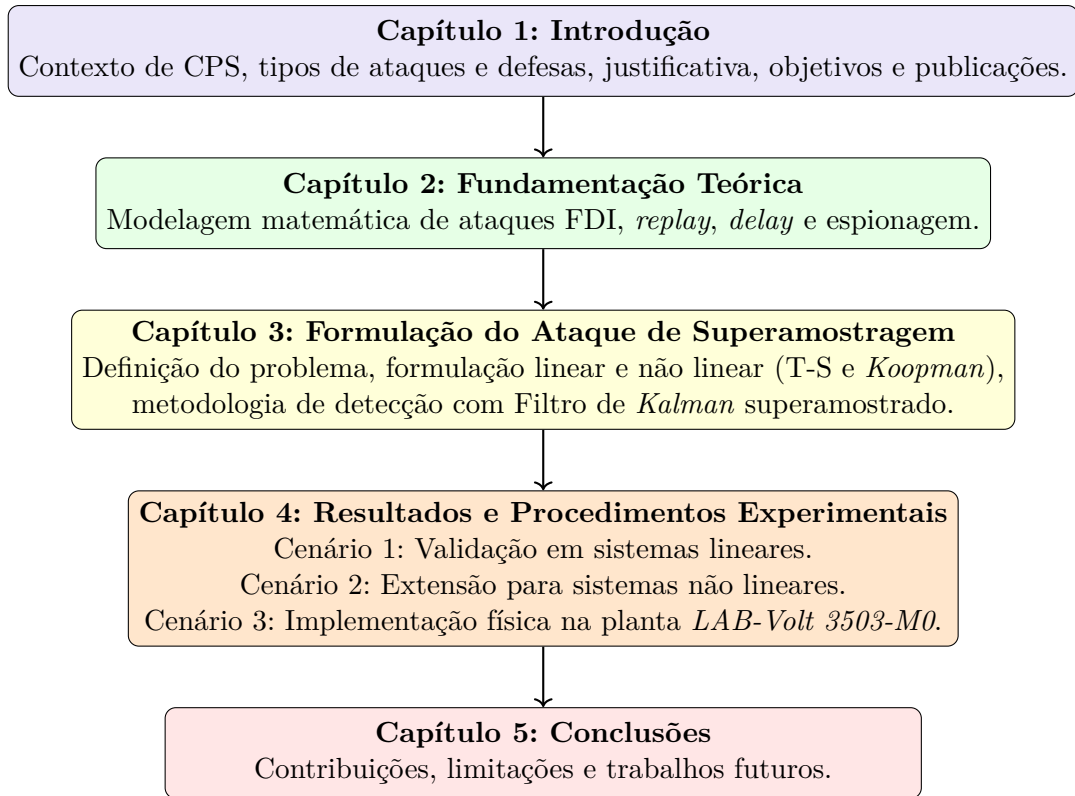


Figura 1 – Esquema geral da organização dos capítulos da dissertação.

2 Fundamentação Teórica

Este capítulo apresenta a fundamentação teórica dos principais tipos de ataque em sistemas ciberfísicos abordados nesta dissertação: injeção de dados falsos, repetição, atraso temporal e espionagem. Para cada tipo, são discutidos o modelo matemático, as condições de indetectabilidade e os métodos de detecção correspondentes.

2.1 Ataques de Injeção de Dados Falsos

Os ataques de injeção de dados falsos constituem uma das ameaças mais estudadas em segurança de CPS, por sua capacidade de comprometer a estimação de estados sem disparar os detectores convencionais. As subseções a seguir detalham sua caracterização, as condições de indetectabilidade e os métodos de detecção disponíveis.

2.1.1 Caracterização do Ataque de Injeção de Dados Falsos

Os ataques de injeção de dados falsos representam uma classe de ameaças ciberfísicas que exploram vulnerabilidades na cadeia de medição e estimação de estados em sistemas de controle industrial. Diferentemente de ataques convencionais que simplesmente interrompem comunicações ou corrompem dados, os ataques FDI são projetados para manipular as informações transmitidas pelos sensores de forma a permanecerem indetectáveis pelos mecanismos tradicionais de detecção baseados em análise de resíduos (LIU; NING; REITER, 2011).

A natureza desse tipo de ataque reside na capacidade do adversário de construir vetores de dados maliciosos que satisfazem as relações algébricas esperadas pelo sistema de estimação de estados. Quando executado adequadamente, o ataque induz o sistema a operar com base em uma percepção falsa da realidade física, potencialmente levando a decisões de controle inadequadas ou prejudiciais (MO; CHABUKSWAR; SINOPOLI, 2013).

Arquitetura do Sistema e Pontos de Injeção

Para compreender o funcionamento do ataque FDI, é necessário primeiro estabelecer a arquitetura típica de um sistema de controle industrial. A Figura 2 ilustra os componentes fundamentais da malha de controle e os pontos vulneráveis onde o ataque pode ser inserido.

O ataque pode ser executado em diferentes pontos da arquitetura. No nível dos sensores, o adversário pode comprometer diretamente os dispositivos de medição, alterando as leituras antes mesmo de sua transmissão. E no canal de comunicação, o atacante

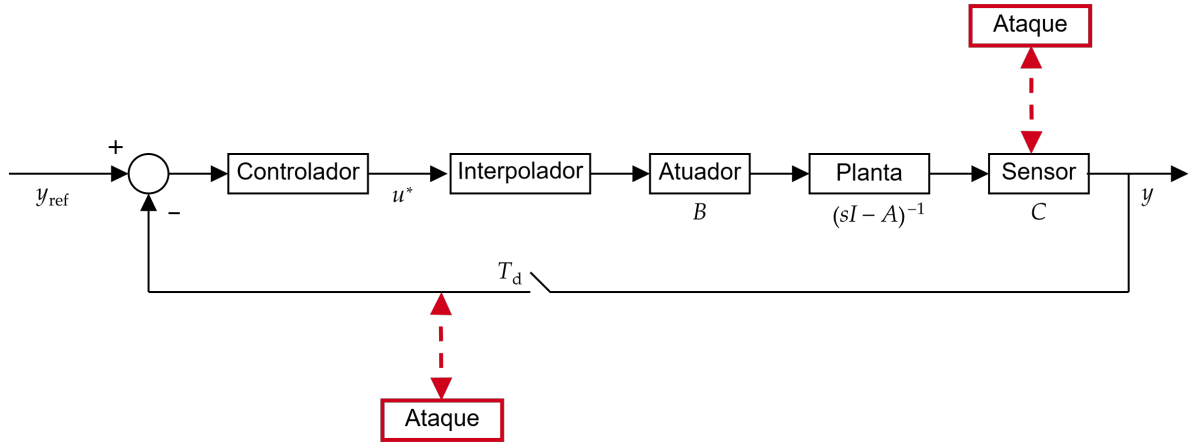


Figura 2 – Arquitetura de sistema de controle sob ataque FDI, mostrando os pontos de inserção do ataque.

intercepta os dados em trânsito e os modifica antes que alcancem o controlador (LIANG et al., 2017).

Modelo Matemático

Considere um sistema de controle industrial modelado em espaço de estados por equações lineares discretas no tempo:

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{b}_w\mathbf{w}(k) \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{v}(k), \end{aligned} \quad (2.1)$$

em que $\mathbf{x}(k) \in \mathbb{R}^n$ representa o vetor de estados do sistema no instante k , $\mathbf{u}(k) \in \mathbb{R}^m$ é o vetor de entradas de controle, $\mathbf{y}(k) \in \mathbb{R}^p$ denota as medições dos sensores, $\mathbf{A} \in \mathbb{R}^{n \times n}$ é a matriz dinâmica, $\mathbf{B} \in \mathbb{R}^{n \times m}$ é a matriz de entrada, e $\mathbf{C} \in \mathbb{R}^{p \times n}$ é a matriz de saída. O termo $\mathbf{b}_w \in \mathbb{R}^n$ representa um vetor de perturbação de processo constante (ou variante no tempo, mas determinístico), que modela desvios na dinâmica do sistema não capturados pela linearização ou entradas de distúrbio conhecidas. Os termos $\mathbf{w}(k)$ e $\mathbf{v}(k)$ representam ruído de processo e de medição, respectivamente, tipicamente modelados como processos estocásticos gaussianos de média nula.

Em condições normais de operação, o sistema de supervisão recebe as medições $\mathbf{z}(k) = \mathbf{y}(k)$ e utiliza um estimador de estados para reconstruir o vetor de estados $\hat{\mathbf{x}}(k)$ a partir das observações disponíveis e do modelo. O estimador mais comum em aplicações industriais é baseado na abordagem de mínimos quadrados ponderados (WLS, do inglês *Weighted Least Squares*), muito utilizado em sistemas de gerenciamento de energia elétrica (KOSUT et al., 2011).

Construção do Vetor de Ataque

Durante um ataque FDI, o adversário injeta um vetor de dados malicioso $\mathbf{a}(k) \in \mathbb{R}^p$ nas medições, de modo que o sistema de supervisão recebe:

$$\mathbf{z}(k) = \mathbf{y}(k) + \mathbf{a}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{v}(k) + \mathbf{a}(k). \quad (2.2)$$

O objetivo do atacante é escolher o vetor $\mathbf{a}(k)$ de forma que as medições corrompidas $\mathbf{z}(k)$ pareçam consistentes com o modelo físico do sistema. Para sistemas lineares com estimação baseada em WLS, o resíduo de detecção é calculado como:

$$\mathbf{r}(k) = \mathbf{z}(k) - \mathbf{C}\hat{\mathbf{x}}(k), \quad (2.3)$$

em que $\hat{\mathbf{x}}(k)$ é a estimativa de estados obtida pela minimização do critério de mínimos quadrados ponderados:

$$\hat{\mathbf{x}}(k) = \arg \min_x [\mathbf{z}(k) - \mathbf{C}\mathbf{x}(k)]^T \mathbf{W} [\mathbf{z}(k) - \mathbf{C}\mathbf{x}(k)], \quad (2.4)$$

sendo $\mathbf{W} \in \mathbb{R}^{p \times p}$ uma matriz de ponderação diagonal positiva definida, geralmente construída a partir das variâncias dos ruídos de medição.

A condição fundamental para a indetectabilidade do ataque foi estabelecida por Liu, Ning e Reiter (2011) e requer que o vetor de ataque satisfaça a relação:

$$\mathbf{a}(k) = \mathbf{C}\mathbf{c}(k), \quad (2.5)$$

em que $\mathbf{c}(k) \in \mathbb{R}^n$ é um vetor escolhido pelo atacante no espaço de estados. Essa condição assegura que o vetor de ataque pertença ao espaço das colunas da matriz de observação $\mathbf{C}$, garantindo que o resíduo pós-ataque mantenha as mesmas propriedades estatísticas do resíduo em condições normais.

2.1.2 Análise de Indetectabilidade

A capacidade de um ataque FDI permanecer indetectável está relacionada à estrutura algébrica do sistema de estimação. Quando a condição expressa na Equação (2.5) é satisfeita, o estado estimado torna-se:

$$\hat{\mathbf{x}}_a(k) = \hat{\mathbf{x}}(k) + \mathbf{c}(k), \quad (2.6)$$

em que $\hat{\mathbf{x}}_a(k)$ indica o estado estimado sob ataque. Note que o mesmo vetor $\mathbf{c}(k)$ utilizado na construção do ataque mostrado na Equação (2.5) aparece agora como viés na estimativa de estados. Consequentemente, o resíduo de detecção torna-se:

$$\begin{aligned} \mathbf{r}_a(k) &= \mathbf{z}(k) - \mathbf{C}\hat{\mathbf{x}}_a(k) \\ &= (\mathbf{C}\mathbf{x}(k) + \mathbf{v}(k) + \mathbf{C}\mathbf{c}(k)) - \mathbf{C}(\hat{\mathbf{x}}(k) + \mathbf{c}(k)) \\ &= \mathbf{C}\mathbf{x}(k) + \mathbf{v}(k) - \mathbf{C}\hat{\mathbf{x}}(k) \\ &= \mathbf{r}(k). \end{aligned} \quad (2.7)$$

Essa derivação demonstra matematicamente que o resíduo sob ataque $\mathbf{r}_a(k)$ é idêntico ao resíduo em condições normais $\mathbf{r}(k)$, tornando o ataque invisível para detectores baseados em análise de resíduos. A magnitude do resíduo permanece estatisticamente consistente com as flutuações esperadas devido ao ruído de medição, não ativando alarmes em sistemas de detecção tradicionais (URBINA et al., 2016).

A indetectabilidade também está relacionada à observabilidade do sistema. Para sistemas completamente observáveis, onde a matriz de observabilidade tem posto completo, o adversário necessita conhecer a matriz $\mathbf{C}$ e ter capacidade de corromper um subconjunto específico de medições. Em sistemas parcialmente observáveis, existem mais graus de liberdade para construção de ataques indetectáveis, aumentando a superfície de vulnerabilidade (MO; CHABUKSWAR; SINOPOLI, 2013).

2.1.3 Métodos de Detecção de Ataques FDI

Embora os ataques FDI possam ser construídos para serem indetectáveis por métodos convencionais baseados em resíduos, diversas abordagens alternativas têm sido desenvolvidas para identificar sua presença. Estas técnicas exploram características que os atacantes não conseguem perfeitamente simular ou que requerem recursos computacionais e informacionais além do que tipicamente está disponível ao adversário.

Detecção Baseada em Dinâmica Física

Uma classe de métodos de detecção explora as restrições impostas pelas leis físicas que governam o sistema. Mesmo quando um ataque satisfaz as relações algébricas estáticas do modelo de estimação, ele pode violar restrições dinâmicas ou limites físicos inerentes ao processo. Por exemplo, em sistemas elétricos, mudanças abruptas no estado estimado que violem as capacidades de rampa dos geradores ou os limites de fluxo de potência podem indicar a presença de dados corrompidos (LIANG et al., 2017).

Métodos baseados em modelos dinâmicos comparam a trajetória estimada com previsões obtidas por meio de simulações físicas do sistema. Inconsistências significativas entre a evolução temporal esperada e a estimativa corrente sugerem a presença de manipulação de dados (KOSUT et al., 2011).

Técnicas Estatísticas

Abordagens estatísticas mais sofisticadas exploram propriedades de segunda ordem e correlações temporais que não são capturadas por testes simples de resíduos. Métodos baseados em análise de componentes principais (PCA, do inglês *principal component analysis*) constroem modelos estatísticos do comportamento normal do sistema e detectam desvios que indicam anomalias. A vantagem dessa abordagem é que ela não requer

conhecimento explícito da matriz $\mathbf{C}$, sendo robusta a incertezas de modelagem (SHARMA et al., 2023).

Técnicas de detecção sequencial, como o teste sequencial de razão de probabilidades (SPRT, do inglês *sequential probability ratio test*), analisam sequências temporais de medições para identificar mudanças sutis nas propriedades estatísticas dos dados. Essas técnicas são particularmente eficazes contra ataques persistentes que gradualmente modificam o estado estimado ao longo do tempo (MO; CHABUKSWAR; SINOPOLI, 2013).

Métodos Baseados em Aprendizado de Máquina

Técnicas de aprendizado de máquina têm sido aplicadas para detecção de ataque FDI, explorando a capacidade de algoritmos de identificar padrões complexos que diferenciam operação normal de cenários de ataque. Redes neurais profundas, como redes neurais recorrentes de memória longa-curta (LSTM, do inglês *long short-term memory*, uma arquitetura de rede neural projetada para capturar dependências temporais de longo prazo em dados sequenciais), têm demonstrado êxito na detecção de ataques ao capturar correlações temporais e espaciais nos dados de medição que se desviam do comportamento esperado (SAYGHE, 2025). Outros trabalhos empregam redes neurais convolucionais em grafos para explorar a topologia da rede e melhorar a localização de ataques (ALRSLANI et al., 2025).

Métodos baseados em máquinas de aprendizado extremo (ELM, do inglês *extreme learning machine*) e *autoencoders* também têm sido propostos para classificação de dados falsos (HU et al., 2023). A principal vantagem dessas abordagens é a capacidade de aprender representações não lineares das relações entre variáveis do sistema, potencialmente capturando assinaturas de ataques que não são evidentes por meio de análise linear. Técnicas de *ensemble* (combinação de múltiplos modelos de aprendizado), como *Random Forest* e *XGBoost*, combinadas com técnicas de balanceamento de dados como SMOTE (gerador de dados sintéticos para balancear classes desbalanceadas), têm alcançado taxas de detecção superiores a 99% em *datasets* de SCADA (DOKKU et al., 2025).

Porém, a eficácia desses métodos depende muito da qualidade e representatividade dos dados de treinamento, e existe o risco de que atacantes criem ataques especialmente projetados para enganar os classificadores.

Proteção Utilizando Medições Seguras

Uma estratégia complementar de defesa envolve a proteção criptográfica de um subconjunto crítico de medições. Se determinadas medições puderem ser asseguradas contra corrupção por meio de canais de comunicação dedicados e autenticação criptográfica, a capacidade do adversário de construir ataques indetectáveis fica reduzida (KIM; POOR,

2011). A escolha ideal desse subconjunto de medições protegidas é um problema de otimização que busca equilibrar o custo de proteção com a cobertura de segurança.

Urbina et al. (2016) demonstraram que mesmo um número relativamente pequeno de medições protegidas pode limitar o impacto de ataques FDI, forçando o adversário a comprometer múltiplos pontos adicionais para manter a indetectabilidade, aumentando assim a complexidade e o risco de detecção do ataque.

2.2 Ataques de Repetição

Os ataques de repetição exploram a consistência temporal dos dados legítimos para enganar o sistema de monitoramento. A seguir, são apresentados sua caracterização, análise de indetectabilidade e as principais estratégias de detecção.

2.2.1 Caracterização do Ataque de Repetição

Os ataques de repetição (*replay attacks*) representam uma classe importante de ameaças ciberfísicas que exploram a natureza temporal dos sistemas de controle. Nesse tipo de ataque, o adversário intercepta e armazena dados legítimos transmitidos pelos sensores durante um período de operação normal do sistema, e posteriormente retransmite essas informações gravadas como se fossem medições em tempo real (MO; SINOPOLI, 2009). A simplicidade conceitual desse ataque contrasta com seu potencial de causar danos significativos ao sistema controlado.

A característica principal dos ataques de repetição é que eles utilizam dados reais do sistema, capturados em um momento anterior, o que torna sua detecção complicada. Como os dados repetidos são internamente consistentes e refletem o comportamento físico válido do sistema no momento da captura, eles podem passar despercebidos por detectores baseados em verificação de consistência estática. O atacante não precisa conhecer o modelo detalhado do sistema nem possuir capacidade de gerar dados sinteticamente coerentes, bastando ter acesso ao canal de comunicação entre sensores e o controlador (ZHU; MARTINEZ, 2013).

Arquitetura do Sistema e Pontos de Injeção

A Figura 3 ilustra a arquitetura de um sistema de controle industrial sob ataque de repetição, destacando os pontos onde o adversário pode interceptar e reinjetar os dados gravados.

Durante a fase de reconhecimento, o atacante monitora passivamente o canal e armazena sequências de medições correspondentes ao comportamento normal do sistema. Posteriormente, durante a fase de ataque ativa, o adversário bloqueia a transmissão das

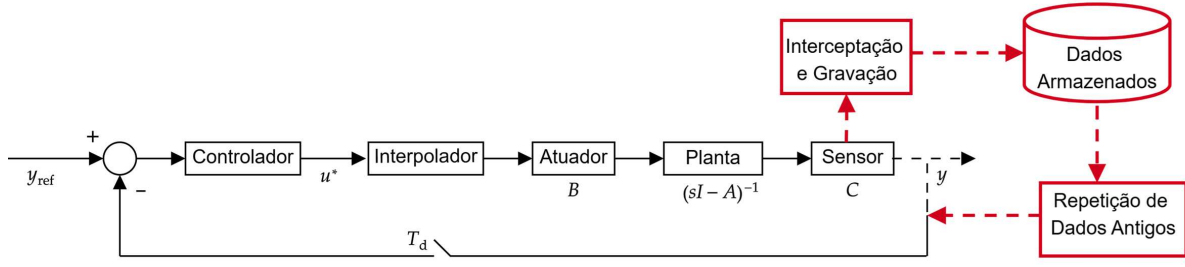


Figura 3 – Arquitetura de sistema de controle sob ataque de repetição, mostrando a interceptação, armazenamento e reinjeção de dados.

medições reais e injeta no seu lugar as medições previamente gravadas. O controlador, ao receber esses dados desatualizados, toma decisões baseadas em uma visão passada do estado real da planta física.

Modelo Matemático do Sistema

Para a construção do ataque, considere o sistema de controle em tempo discreto descrito pela Equação (2.1).

Construção do Ataque de Repetição

Durante o ataque de repetição, o adversário substitui as medições atuais $\mathbf{y}(k)$ por medições previamente gravadas $\mathbf{y}_r(k - \tau)$, sendo $\tau > 0$ o atraso temporal entre o momento da gravação e o momento da repetição. Formalmente, as medições recebidas pelo estimador durante o ataque são:

$$\mathbf{z}(k) = \begin{cases} \mathbf{y}(k) & \text{se não há ataque} \\ \mathbf{y}_r(k - \tau) & \text{se há ataque.} \end{cases} \quad (2.8)$$

O conjunto de dados repetidos $\{\mathbf{y}_r(k - \tau)\}$ foi capturado durante o intervalo $[k_s, k_e]$, sendo k_s e k_e denotam os instantes de início e fim da gravação, respectivamente. Durante esse período de captura, o sistema operava em condições normais, portanto:

$$\mathbf{y}_r(k - \tau) = \mathbf{C}\mathbf{x}(k - \tau) + \mathbf{v}(k - \tau), \quad k \in [k_s + \tau, k_e + \tau]. \quad (2.9)$$

Uma característica importante é que o atacante pode repetir os dados múltiplas vezes ou em padrões não sequenciais, aumentando a flexibilidade do ataque. O adversário também pode combinar segmentos de diferentes períodos de gravação para criar trajetórias específicas de comportamento do sistema controlado.

2.2.2 Análise de Indetectabilidade

Considere um detector de anomalias baseado na análise de inovações do filtro de *Kalman*. O resíduo da sequência é definido como:

$$\mathbf{r}(k) = \mathbf{z}(k) - \mathbf{C}\hat{\mathbf{x}}(k|k-1), \quad (2.10)$$

em que $\hat{\mathbf{x}}(k|k-1)$ é a predição do estado baseada nas informações até o instante $k-1$. Em operação normal, o resíduo $\mathbf{r}(k)$ é um processo com média nula e covariância conhecida $\mathbf{S}(k)$.

Durante um ataque de repetição, o resíduo torna-se:

$$\mathbf{r}_a(k) = \mathbf{y}_r(k-\tau) - \mathbf{C}\hat{\mathbf{x}}(k|k-1). \quad (2.11)$$

A propriedade que permite a indetectabilidade temporária é que, se o sistema for estável e o controlador mantiver o sistema próximo a um ponto de operação, as medições repetidas $\mathbf{y}_r(k-\tau)$ podem permanecer consistentes com as predições do estado por um período finito. Especificamente, se o sistema estava em regime permanente durante a gravação e continua em regime permanente durante a repetição, os resíduos podem não exceder significativamente os limiares de detecção (FANG et al., 2020).

No entanto, existe uma diferença fundamental: as medições repetidas não incorporam as mudanças no estado real da planta causadas pelas perturbações $w(k)$ ou pelas variações no controle $\mathbf{u}(k)$. Com o tempo, essa discrepância se acumula, levando a resíduos crescentes. Para sistemas instáveis (com autovalores de $\mathbf{A}$ fora do círculo unitário), a divergência é exponencial (SANCHEZ et al., 2019).

A janela de indetectabilidade T_d pode ser caracterizada pela condição:

$$\|\mathbf{r}_a(k)\| < \gamma, \quad \forall k \in [k_a, k_a + T_d], \quad (2.12)$$

em que γ é o limiar de detecção e k_a é o instante de início do ataque. O valor de T_d depende da estabilidade do sistema, da magnitude das perturbações e da qualidade das medições gravadas.

2.2.3 Métodos de Detecção de Ataques de Repetição

Métodos Baseados em Marcação d'Água Física

Uma das abordagens mais utilizadas para detecção de ataques de repetição envolve a técnica de marcação d'água física (*physical watermarking*). Nesse método, um sinal de excitação conhecido é intencionalmente adicionado à entrada de controle do sistema. Esse sinal funciona como uma marca d'água que deve estar presente nas medições subsequentes dos sensores (MO; SINOPOLI, 2009).

Formalmente, a entrada de controle (assumindo que é por realimentação de estados) é modificada para:

$$\mathbf{u}(k) = -\mathbf{K}\hat{\mathbf{x}}(k) + \mathbf{e}(k), \quad (2.13)$$

em que $\mathbf{e}(k)$ é o sinal de marcação d'água, tipicamente escolhido como um processo gaussiano de média nula com covariância Σ_e . O sinal de marcação deve ser suficientemente pequeno para não comprometer significativamente o desempenho do controlador, mas grande o suficiente para permitir detecção confiável.

Durante a operação normal, o sinal de marcação d'água presente na entrada de controle se propaga utilizando a dinâmica do sistema e aparece nas medições dos sensores. A correlação entre o sinal injetado $\mathbf{e}(k)$ e as medições recebidas $\mathbf{z}(k)$ é calculada:

$$\rho(k) = \frac{1}{W} \sum_{i=k-W+1}^k \mathbf{e}^T(i)\mathbf{z}(i), \quad (2.14)$$

em que W é o tamanho da janela de correlação. Durante um ataque de repetição, as medições retransmitidas não contêm o sinal de marcação d'água atual, resultando em uma queda na correlação $\rho(k)$. O ataque é detectado quando $\rho(k) < \rho_{th}$, em que ρ_{th} é um limiar apropriadamente escolhido.

Detecção Baseada em Análise de Frequência

Métodos baseados em análise de frequência exploram a assinatura espectral das medições para identificar padrões repetitivos. Em sistemas dinâmicos reais, pequenas variações estocásticas nas perturbações e ruídos de medição garantem que as medições nunca sejam exatamente idênticas ao longo do tempo. Durante um ataque de repetição, a sequência retransmitida exibe periodicidade ou padrões de repetição detectáveis por meio de técnicas de processamento de sinais (SANCHEZ et al., 2019).

Uma abordagem utiliza a transformada discreta de Fourier (DFT, do inglês *discrete Fourier transform*) das medições para identificar componentes de frequência anormalmente fortes que indicam periodicidade artificial (SANCHEZ et al., 2019). Outra técnica emprega a entropia espectral para quantificar a aleatoriedade das medições, onde valores reduzidos de entropia sugerem a presença de padrões repetitivos inconsistentes (ZHOU; ZHANG; XIE, 2021).

Métodos Baseados em Estimação Dinâmica de Atrasos

Técnicas propõem a detecção de ataques de repetição utilizando a estimação da defasagem temporal entre as medições recebidas e o estado real do sistema. O método caracteriza a flutuação dinâmica do atraso aparente entre as medições e as previsões do modelo (ZHU; MARTINEZ, 2013).

A lógica de detecção compara as características estatísticas da variação de atraso durante operação normal com aquelas observadas quando dados antigos estão sendo repetidos. Durante um ataque, a correlação temporal entre medições e estados estimados exibe padrões anômalos que revelam a presença de dados desatualizados.

Detecção por Análise de Covariância

Métodos estatísticos exploram mudanças na matriz de covariância dos resíduos de inovação. Durante operação normal, a covariância dos resíduos segue propriedades conhecidas derivadas do filtro de *Kalman*. Sob ataque de repetição, a covariância observada dos resíduos diverge daquela esperada teoricamente (FRANZÈ; TEDESCO; FAMULARO, 2020).

A detecção é implementada utilizando testes estatísticos como a divergência de *Kullback-Leibler* entre a distribuição esperada e observada dos resíduos. Quando a divergência excede um limiar, o sistema identifica a presença de um ataque. Essa abordagem é eficaz porque não requer modificações no controle e opera puramente por meio do monitoramento passivo das estatísticas dos dados.

Detecção por Codificação da Saída

Uma estratégia atual envolve a aplicação de esquemas de codificação nas saídas dos sensores que transformam ataques de repetição naturalmente indetectáveis em ataques aditivos detectáveis. A ideia fundamental é adicionar um código dependente do tempo às medições que varia de forma conhecida pelo sistema legítimo mas imprevisível para o atacante (GUO et al., 2021).

Formalmente, a medição codificada é:

$$\tilde{\mathbf{y}}(k) = \mathbf{y}(k) + \phi(k), \quad (2.15)$$

em que $\phi(k)$ é um código dependente do tempo conhecido pelo estimador. Quando o atacante repete medições codificadas de um período anterior, a codificação desatualizada é detectada pela incompatibilidade entre o código esperado e o código recebido. Essa técnica converte o problema de detecção de repetição temporal em um problema de detecção de anomalia de codificação.

2.3 Ataques de Atraso Temporal

Os ataques de atraso temporal atuam sobre as características temporais das comunicações, comprometendo a estabilidade e o desempenho do controlador sem necessariamente alterar o conteúdo dos dados. As subseções seguintes descrevem sua formulação, as condições de indetectabilidade e os métodos de detecção associados.

2.3.1 Caracterização do Ataque de Atraso Temporal

Os ataques de atraso temporal (*time-delay attacks*) representam uma classe de ameaças ciberfísicas que manipulam as características temporais dos canais de comunicação em sistemas de controle. Diferentemente dos ataques de repetição que retransmitem dados antigos, ou dos ataques de injeção de dados falsos que modificam o conteúdo das mensagens, os ataques de atraso temporal interferem com o momento em que a informação é entregue, introduzindo atrasos artificiais na transmissão de dados entre os componentes do sistema (BIANCHIN; PASQUALETTI, 2018).

A motivação desse tipo de ataque reside no fato de que muitos sistemas de controle são projetados assumindo comunicações próximas ao tempo real, e sua estabilidade e desempenho dependem da entrega oportuna de informações. Ao introduzir atrasos além do que o sistema foi projetado para tolerar, o atacante pode degradar o desempenho do controlador ou até mesmo levar o sistema à instabilidade, sem necessariamente modificar o conteúdo dos dados transmitidos (SARGOLZAEI; YEN; ABDELGHANI, 2015).

A execução de ataques de atraso temporal pode ocorrer de diversas formas. O adversário pode atuar como um intermediário nos canais de comunicação, retendo pacotes por períodos variáveis antes de enviá-los ao destino. Outra forma é o atacante criar congestionamento artificial na rede utilizando ataques de negação de serviço coordenados que aumentam o tempo de transmissão de forma seletiva. Uma característica desses ataques é que eles podem ser difíceis de distinguir de atrasos normais causados por congestionamento natural da rede ou falhas temporárias de comunicação.

Arquitetura do Sistema e Pontos de Injeção

A Figura 4 ilustra a arquitetura de um sistema de controle industrial sob ataque de atraso temporal, destacando os canais de comunicação vulneráveis onde atrasos artificiais podem ser inseridos.

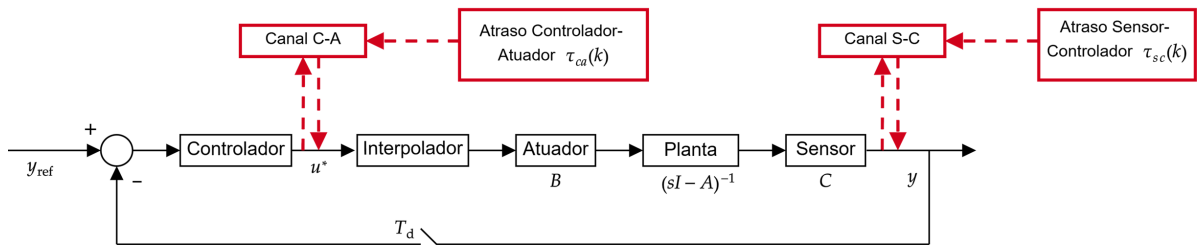


Figura 4 – Arquitetura de sistema de controle sob ataque de atraso temporal, mostrando os canais vulneráveis sensor-controlador (S-C) e controlador-atuador (C-A).

Os ataques de atraso temporal podem ser inseridos em dois locais principais: no canal sensor-controlador (S-C) e no canal controlador-atuador (C-A). No primeiro caso, as medições dos sensores são atrasadas antes de chegarem ao controlador, resultando em

decisões de controle baseadas em informações desatualizadas sobre o estado da planta. No segundo caso, os comandos de controle são atrasados antes de alcançarem os atuadores, impedindo que a ação corretiva seja aplicada no momento adequado.

Modelo Matemático do Sistema

Considere um sistema de controle em tempo discreto descrito pelas equações de espaço de estados:

$$\begin{aligned}\mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k - \tau_{ca}(k)) + \mathbf{b}_w w(k) \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{v}(k),\end{aligned}\tag{2.16}$$

em que $\tau_{ca}(k)$ denota o atraso no canal controlador-atuador.

O controlador (obtido por realimentação de estados) recebe medições atrasadas e é descrito por:

$$\mathbf{u}(k) = -\mathbf{K}\mathbf{z}(k - \tau_{sc}(k)),\tag{2.17}$$

em que $\mathbf{z}(k - \tau_{sc}(k)) = \mathbf{y}(k - \tau_{sc}(k))$ representa a medição atrasada devido ao canal S-C, $\tau_{sc}(k)$ é o atraso nesse canal, e $\mathbf{K}$ é a matriz de ganho de realimentação.

Substituindo a Equação (2.17) na Equação (2.16), o sistema em malha fechada sob ataque de atraso pode ser escrito como:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) - \mathbf{B}\mathbf{K}\mathbf{y}(k - \tau_{sc}(k) - \tau_{ca}(k)) + \mathbf{b}_w w(k).\tag{2.18}$$

Essa formulação mostra que o atraso total efetivo experimentado pelo sistema é a soma dos atrasos em ambos os canais: $\tau_{total}(k) = \tau_{sc}(k) + \tau_{ca}(k)$.

Construção do Ataque de Atraso Temporal

Durante um ataque de atraso temporal, o adversário manipula os atrasos nos canais de comunicação de forma intencional. Os atrasos podem ser modelados como:

$$\begin{aligned}\tau_{sc}(k) &= \tau_{sc}^{nom} + \Delta\tau_{sc}^{atk}(k) \\ \tau_{ca}(k) &= \tau_{ca}^{nom} + \Delta\tau_{ca}^{atk}(k),\end{aligned}\tag{2.19}$$

em que τ_{sc}^{nom} e τ_{ca}^{nom} representam os atrasos nominais inerentes ao sistema de comunicação, e $\Delta\tau_{sc}^{atk}(k)$ e $\Delta\tau_{ca}^{atk}(k)$ denotam os atrasos adicionais introduzidos pelo atacante.

O atacante pode adotar diferentes estratégias na escolha dos atrasos falsos. Uma estratégia simples consiste em aplicar atrasos constantes suficientemente grandes para degradar o desempenho. Alternativamente, o adversário pode usar atrasos variáveis no tempo, escolhidos de forma a maximizar o impacto sobre o sistema. Para sistemas cuja estabilidade depende de margens de atraso conhecidas, o atacante pode calibrar $\Delta\tau^{atk}(k)$ para levar o atraso total além do valor crítico de estabilidade.

Uma formulação mais geral considera atrasos dependentes do estado ou do tempo (SARGOLZAEI et al., 2017):

$$\Delta\tau^{atk}(k) = f(\mathbf{x}(k), k), \quad (2.20)$$

em que $f(\cdot)$ é uma função escolhida pelo atacante para maximizar alguma métrica de impacto, como o erro de rastreamento ou a degradação de desempenho do controlador.

2.3.2 Análise de Indetectabilidade

A indetectabilidade dos ataques de atraso temporal está na dificuldade de distinguir atrasos maliciosos de variações legítimas na latência da rede. Em sistemas práticos, os canais de comunicação exibem naturalmente variações estocásticas de atraso devido a congestionamento, roteamento dinâmico, e condições da rede física.

Considere que os atrasos nominais do sistema seguem uma distribuição de probabilidade conhecida $p_{nom}(\tau)$. Durante um ataque, a distribuição observada torna-se $p_{atk}(\tau)$. O atacante pode construir atrasos maliciosos que satisfazem:

$$D_{KL}(p_{atk}||p_{nom}) < \epsilon, \quad (2.21)$$

em que D_{KL} denota a divergência de *Kullback-Leibler* e ϵ é um limiar pequeno. Essa condição garante que a distribuição de atrasos sob ataque seja estatisticamente próxima à distribuição nominal, tornando a detecção baseada em análise estatística simples ineficaz.

A margem de estabilidade do sistema em relação a atrasos pode ser caracterizada por um atraso máximo admissível τ_{max} tal que o sistema permanece estável para $\tau_{total}(k) \leq \tau_{max}$. Esse valor crítico pode ser determinado por meio de análise de *Lyapunov-Krasovskii* para sistemas com atraso (LOU et al., 2019).

O atacante escolhe atrasos que mantêm o sistema estável por períodos prolongados (evitando detecção imediata por falhas catastróficas), mas gradualmente degrada o desempenho. Essa degradação pode manifestar-se como aumento no tempo de acomodação, *overshoot* maior, ou erro de regime permanente elevado, características que podem ser atribuídas a outros fatores operacionais.

2.3.3 Métodos de Detecção de Ataques de Atraso Temporal

Detecção Baseada em Análise Estatística de Atrasos

Uma abordagem para detectar ataques de atraso temporal envolve o monitoramento contínuo das estatísticas dos atrasos de comunicação e a identificação de desvios em relação aos padrões esperados. O método estima a distribuição dos atrasos observados e compara com um modelo de referência estabelecido durante operação normal (BIANCHIN; PASQUALETTI, 2018).

Seja $\{\tau_1, \tau_2, \dots, \tau_N\}$ uma janela temporal de atrasos observados. Estatísticas de primeira e segunda ordem, como média $\bar{\tau}$ e variância σ_τ^2 , são calculadas. A detecção é baseada em testes de hipótese:

$$\mathbf{H}_0 : \bar{\tau} \in [\mu_0 - \delta, \mu_0 + \delta] \quad \text{vs} \quad \mathbf{H}_1 : \bar{\tau} \notin [\mu_0 - \delta, \mu_0 + \delta], \quad (2.22)$$

em que μ_0 é a média nominal esperada e δ é uma margem de tolerância baseada nas características convencionais da rede.

Técnicas mais sofisticadas empregam testes de razão de verossimilhança para distinguir entre distribuições de atraso nominais e aquelas induzidas por ataques, explorando mudanças nas características de probabilidade dos tempos de chegada de pacotes (ZHANG et al., 2022).

Detecção Baseada em Análise de Desempenho de Controle

Métodos alternativos detectam ataques de atraso temporal mediante o monitoramento do desempenho do sistema de controle. Métricas de desempenho como o erro de rastreamento, a norma do sinal de controle, e o tempo de acomodação são continuamente avaliadas (LOU et al., 2019).

Define-se um índice de desempenho baseado no erro quadrático:

$$J(k) = \sum_{i=k-W}^k \|\mathbf{x}(i) - \mathbf{x}_{ref}(i)\|^2, \quad (2.23)$$

em que W é o tamanho da janela de avaliação e $\mathbf{x}_{ref}$ é a trajetória de referência desejada. Quando $J(k)$ excede os valores históricos sem mudanças correspondentes nas condições operacionais ou perturbações conhecidas, isso pode indicar a presença de atrasos maliciosos.

Essa abordagem é particularmente segura quando combinada com modelos preditivos do comportamento esperado do sistema, permitindo a detecção de degradação anômala de desempenho mesmo quando os atrasos individuais permanecem dentro de faixas aparentemente aceitáveis.

Técnicas Baseadas em Modelos de Aprendizado

Abordagens baseadas em aprendizado de máquina têm sido aplicadas para construir modelos que aprendem os padrões normais de variação de atraso e desempenho do sistema. Redes neurais recorrentes, particularmente LSTM, podem capturar dependências temporais nos padrões de atraso e identificar anomalias que desviam do comportamento aprendido (GANESH et al., 2021).

Um classificador é treinado utilizando dados históricos rotulados, contendo exemplos de operação normal e de cenários de ataque. Durante a operação, sequências de atrasos observados são alimentadas ao modelo, que produz uma probabilidade de ataque. A detecção é acionada quando essa probabilidade excede um limiar predefinido.

Controle Resiliente com Compensação de Atraso

Uma estratégia complementar envolve o projeto de controladores resilientes a atrasos. Técnicas de controle preditivo podem incorporar modelos de atraso e projetar ações de controle que mantêm estabilidade e desempenho aceitáveis mesmo na presença de atrasos variáveis (SARGOLZAEI; YEN; ABDELGHANI, 2015).

Controladores baseados em preditor de *Smith* ou suas variantes podem compensar atrasos conhecidos, embora sua eficácia seja limitada quando os atrasos são altamente variáveis ou desconhecidos. Abordagens adaptativas que estimam os atrasos em tempo real e ajustam os parâmetros do controlador de acordo oferecem maior robustez (INGIMUNDARSON; HÄGGLUND, 2001).

Uma arquitetura de controle resiliente pode implementar múltiplos controladores projetados para diferentes cenários de atraso, com um mecanismo de chaveamento que seleciona o controlador apropriado baseado na estimativa corrente dos atrasos. Esse conceito de controle chaveado com dependência de atraso tem demonstrado capacidade de manter estabilidade e desempenho sob ataques de atraso temporal (SARGOLZAEI et al., 2017).

2.4 Ataques de Espionagem

Os ataques de espionagem comprometem a confidencialidade das comunicações industriais de forma passiva, sem interferir no funcionamento do sistema. A seguir, são apresentados sua caracterização, as razões de sua indetectabilidade e as contramedidas aplicáveis.

2.4.1 Caracterização do Ataque de Espionagem

Os ataques de espionagem (*eavesdropping attacks*) representam uma classe de ameaças ciberfísicas que comprometem a confidencialidade das comunicações em sistemas de controle industrial. Diferentemente dos ataques de injeção de dados falsos, ou dos ataques de atraso temporal, os ataques de espionagem interceptam passivamente o tráfego de rede para obter informações sensíveis sem necessariamente alterar ou interromper as comunicações do sistema (AMIN et al., 2012).

A motivação desse tipo de ataque reside no fato de que muitos sistemas SCADA e de controle industrial foram originalmente projetados para ambientes isolados, sem considerar mecanismos de criptografia ou autenticação. Com a crescente interconexão desses sistemas às redes corporativas e à *Internet*, atacantes podem explorar a falta de criptografia em protocolos industriais antigos como *Modbus*, DNP3 (protocolo de comunicação distribuída amplamente utilizado em automação de energia e saneamento) e IEC 60870-5-104 (padrão internacional para telecontrole de sistemas elétricos via redes TCP/IP) para capturar

dados críticos, incluindo medições de sensores, comandos de controle, credenciais de acesso e informações sobre a topologia e configuração do sistema (SAMTANI et al., 2016).

Ataques de espionagem ocorrem de várias formas. O adversário pode usar a técnica *man-in-the-middle*, colocando-se entre os dispositivos para ler todo o tráfego trocado. Outra opção é a escuta passiva, onde o atacante apenas coleta dados da rede sem interferir nos equipamentos. A principal característica aqui é a dificuldade de detecção, como esses ataques não alteram nem bloqueiam a comunicação, eles deixam poucos vestígios no sistema.

Arquitetura do Sistema e Pontos de Injeção

A Figura 5 ilustra a arquitetura de um sistema de controle industrial sob ataque de espionagem, destacando os pontos vulneráveis onde a interceptação de dados pode ocorrer.

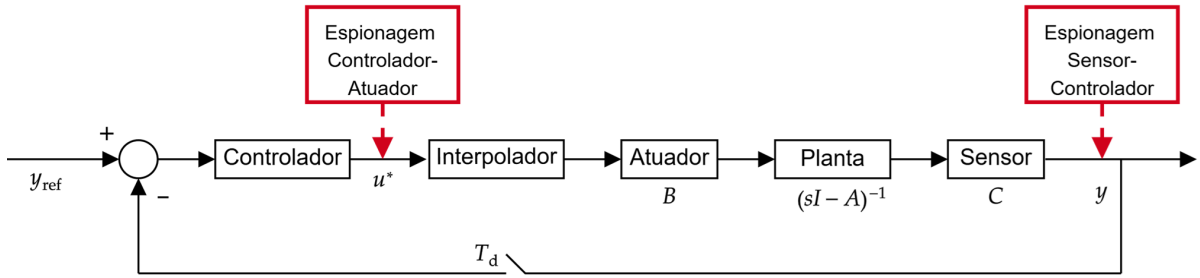


Figura 5 – Arquitetura de sistema de controle sob ataque de espionagem, mostrando os pontos de interceptação nos canais sensor-controlador e controlador-atuador.

Os ataques de espionagem podem ser realizados em múltiplos pontos da arquitetura: entre sensores e controlador, entre controlador e atuadores, ou em qualquer segmento de rede que transporta dados não criptografados. No primeiro caso, o atacante obtém acesso a todas as medições do processo, permitindo compreender o estado operacional do sistema. No segundo caso, os comandos de controle são visíveis, revelando as ações de controle aplicadas e potencialmente os algoritmos de controle utilizados. A interceptação simultânea em ambos os canais fornece ao adversário uma visão completa do comportamento do sistema em malha fechada.

Modelo Matemático do Ataque

Considere o sistema de controle descrito pela Equação (2.1).

Durante um ataque de espionagem, o adversário intercepta sequências de medições e comandos de controle ao longo do tempo. O conjunto de dados interceptados pode ser representado como:

$$\mathcal{D}_{int} = \{(\mathbf{y}(k), \mathbf{u}(k)) : k \in]k_0, k_0 + N]\}, \quad (2.24)$$

em que k_0 é o instante inicial da espionagem e N é o número de amostras capturadas.

2.4.2 Análise de Indetectabilidade

A indetectabilidade dos ataques de espionagem está fundamentada em sua natureza passiva e não intrusiva. Ao contrário de ataques ativos que modificam, atrasam ou bloqueiam comunicações, a espionagem apenas observa o tráfego existente sem introduzir anomalias detectáveis no comportamento do sistema.

A principal característica que torna a detecção desafiadora é que o atacante não precisa injetar pacotes, modificar temporizações, ou alterar padrões de tráfego. O sistema de controle continua operando normalmente, e todas as métricas de desempenho permanecem dentro dos limites esperados.

Do ponto de vista da rede, um ataque de espionagem passivo não gera tráfego adicional detectável. Em segmentos de rede com *hubs* ou em implementações com configurações inadequadas de segmentação, a espionagem passiva pode ocorrer sem deixar rastros.

A probabilidade de detecção de espionagem passiva aumenta quando o atacante precisa realizar ações preparatórias como varredura de rede, ARP (do inglês *address resolution protocol*) *spoofing* (falsificação de endereços MAC (do inglês *media access control*) na rede local para redirecionar o tráfego), ou comprometimento de dispositivos intermediários, que podem gerar *logs* e alertas detectáveis.

2.4.3 Métodos de Detecção de Ataques de Espionagem

Detecção Baseada em Análise de Tráfego de Rede

Embora a espionagem passiva seja difícil de detectar diretamente, técnicas preparatórias utilizadas pelo atacante frequentemente deixam rastros detectáveis. Sistemas de detecção podem monitorar padrões anômalos na camada de enlace e rede (SAMTANI et al., 2016).

Ataques de ARP *spoofing*, comumente usados para posicionar o atacante como intermediário em redes comutadas, podem ser detectados mediante o monitoramento de tabelas ARP:

$$\text{Anomalia} = |\{\text{MAC} : \text{IP} \rightarrow \text{MAC}\} \cap \mathcal{T}_h| < \theta, \quad (2.25)$$

em que $\mathcal{T}_h$ representa mapeamentos históricos IP-MAC conhecidos e θ é um limiar de similaridade.

Sistemas de detecção de intrusão em rede podem identificar tentativas de varredura de rede, acesso não autorizado a portas de comunicação, ou tentativas de exploração de vulnerabilidades conhecidas em protocolos SCADA. Ferramentas como *Snort* (sistema de detecção de intrusões baseado em regras), *Zeek* (plataforma de análise de protocolo e monitoramento de rede), ou soluções especializadas para ambientes industriais (como

Nozomi Networks, *Clarity* e *Dragos*) podem ser configuradas com regras específicas para padrões de tráfego SCADA (LIN et al., 2013).

Detecção Baseada em Monitoramento de Dispositivos

Uma abordagem complementar envolve o monitoramento ativo da integridade e autenticidade dos dispositivos na rede industrial. O inventário ativo de dispositivos mantém um registro atualizado de todos os equipamentos autorizados na rede, com verificação periódica por meio de varreduras controladas, permitindo identificar dispositivos não reconhecidos que podem indicar presença de equipamentos de espionagem ou comprometimento de nós legítimos (YANG et al., 2014). Adicionalmente, a detecção de interfaces operando em modo de captura irrestrita constitui um mecanismo importante de segurança, uma vez que placas de rede nesse modo capturam todo o tráfego do segmento *Ethernet*, não apenas pacotes endereçados a elas. Técnicas especializadas podem detectar esse comportamento por meio de análise de tempos de resposta ou comportamento de pilha de protocolos (YANG et al., 2013).

A análise de fluxo de dados complementa essas técnicas mediante a implementação de contadores e métricas de fluxo em comutadores e roteadores, possibilitando identificar padrões anômalos de tráfego, como volumes inesperados de dados copiados para determinadas portas ou espelhamento não autorizado de tráfego (GOLDENBERG; WOOL, 2013).

Detecção Baseada em Autenticação Mútua e Verificação de Integridade

Protocolos que implementam autenticação mútua e verificação criptográfica de integridade podem dificultar ataques de espionagem, particularmente aqueles que requerem posicionamento ativo como *Man-in-the-Middle*.

A implementação de códigos de autenticação de mensagem ou assinaturas digitais permite que receptores verifiquem a autenticidade e integridade das mensagens (SCHLEGEL; OBERMEIER; SCHNEIDER, 2017).

Protocolos como o IEC 62351 (norma de segurança cibernética projetada para adicionar autenticidade, confidencialidade e integridade à série IEC 61850 e outros protocolos de telecontrole) e IEEE 1588 (para protocolo de sincronização de tempo) incorporam mecanismos de segurança que, quando devidamente implementados, reduzem a superfície de ataque disponível para espionagem (HUSSAIN; USTUN; KALAM, 2020; SHEREEN et al., 2019).

Técnicas Baseadas em Aprendizado de Máquina

Abordagens baseadas em aprendizado de máquina podem ser treinadas para identificar padrões sutis que indicam preparação ou execução de ataques de espionagem, incluindo sequências atípicas de acesso a recursos, tentativas de autenticação em horários incomuns, desvios na implementação de protocolos, ou uso de campos opcionais raramente utilizados (PONOMAREV; ATKISON, 2016). Classificadores baseados em árvores de decisão, *Random Forests* ou redes neurais profundas podem ser treinados com dados rotulados de operação normal e tentativas conhecidas de espionagem, permitindo que o monitoramento contínuo de características extraídas do tráfego de rede produza alertas quando comportamentos suspeitos são identificados (MAGLARAS; JIANG, 2014).

Técnicas Baseadas em Criptografia

A contramedida principal contra ataques de espionagem é a implementação de criptografia em todos os canais de comunicação. Sistemas modernos devem utilizar protocolos seguros sempre que possível, proporcionando confidencialidade e integridade dos dados por meio de criptografia simétrica, e *IPsec* (protocolo de segurança de *Internet* que autentica e criptografa pacotes de dados na camada de rede) para segurança na camada de rede, sendo útil para criar redes privadas virtuais (VPN, do inglês *virtual private network*) entre sites distribuídos de sistemas SCADA.

Para dispositivos com recursos computacionais limitados, como sensores IoT (do inglês *Internet of things*) e RTUs (do inglês *remote terminal unit*) de baixo custo, algoritmos de criptografia leve como AES-GCM, que combina criptografia de dados com autenticação de mensagens em uma única operação, oferecem segurança adequada sem sobrecarga computacional (KIM et al., 2020). Abordagens alternativas incluem o uso de sistemas caóticos para mascarar medições de sensores antes da transmissão. Nessa técnica, um sinal caótico é adicionado à medição real, dificultando que atacantes consigam estimar os estados do sistema mesmo interceptando os dados transmitidos (CHEN et al., 2024). A vantagem dessa abordagem é que ela não requer canais de comunicação adicionais para sincronização, e o efeito de mascaramento pode ser removido no estimador remoto utilizando observadores estendidos que estimam simultaneamente os estados do sistema físico e do sistema caótico (CECILIA et al., 2023).

A seleção de mecanismos de criptografia deve considerar o equilíbrio entre segurança e latência. De acordo com Teixeira et al. (2015), em sistemas de controle em tempo real, atrasos introduzidos por operações de criptografias devem ser mantidos abaixo de limiares críticos para garantir a estabilidade do sistema.

3 Formulação do Ataque de Superamostragem e Metodologia de Detecção

Este capítulo apresenta a formulação matemática do ataque de superamostragem e a metodologia proposta para sua detecção. Inicialmente, define-se o problema e descreve-se a arquitetura do ataque; em seguida, desenvolvem-se as formulações para sistemas lineares e não lineares; e, por fim, apresenta-se a estratégia de detecção baseada no filtro de *Kalman* superamostrado.

3.1 Definição do Problema

O ataque de superamostragem representa uma nova classe de ameaça aos sistemas de controle amostrados, explorando a diferença entre as amostras para comprometer a operação do sistema sem ser detectado pelos métodos convencionais de monitoramento. Esse ataque manipula a ação de controle em uma frequência mais alta que a original, mantendo a variável controlada inalterada nos instantes de amostragem normais.

Considere um sistema de controle digital operando com período de amostragem T_D . A lei de controle é projetada e implementada assumindo esse tempo de amostragem específico, e a ação de controle é enviada ao atuador. A Figura 6 apresenta o diagrama completo do sistema sob ataque, mostrando como a ação de controle original $\mathbf{u}^*$ é interceptada e substituída pela ação gerada pelo sistema superamostrado $\mathbf{u}_a$.

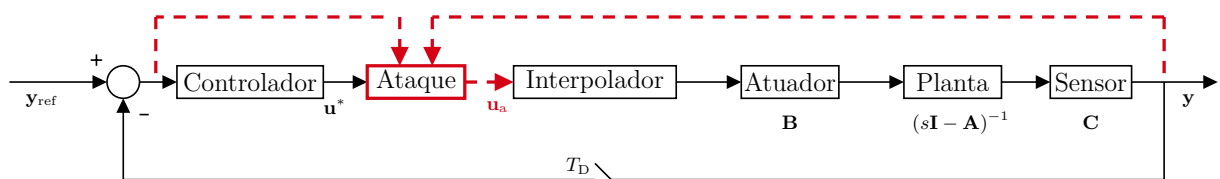


Figura 6 – Diagrama de blocos do sistema de controle sob ataque de superamostragem.

O atacante opera com um período de amostragem menor $T_d < T_D$, obtendo um modelo discretizado da planta com maior resolução temporal. O desafio consiste em produzir uma nova lei de controle que opere com período T_d de tal maneira que a variável controlada permaneça inalterada nos instantes nT_D , $n = 0, 1, 2, \dots$, tornando o ataque difícil de detectar pelos métodos convencionais.

Para que o ataque de superamostragem seja viável, três condições devem ser satisfeitas simultaneamente. Primeiro, o atacante precisa ter acesso ao canal entre o controlador e o atuador, de modo a interceptar o sinal de controle $\mathbf{u}^*$ e substituí-lo pela ação maliciosa $\mathbf{u}_a$. Segundo, o adversário deve dispor de um modelo suficientemente preciso

da planta, a fim de discretizá-la com período T_d e calcular, a cada instante, as ações de controle que satisfazem a condição de coincidência nos instantes de amostragem originais. Terceiro, o sistema de atuação deve ser capaz de responder à taxa de superamostragem $1/T_d$, ou seja, o atuador não pode ser mais lento do que o período de injeção do ataque. Quando essas três condições são atendidas, o ataque consegue manipular a dinâmica da planta entre as amostras sem deixar rastros detectáveis nos instantes de monitoramento do controlador.

3.2 Formulação do Ataque em Sistemas Lineares

Nesta seção, apresenta-se a formulação matemática do ataque de superamostragem para sistemas lineares invariantes no tempo (LTI, do inglês *linear time-invariant*) com múltiplas entradas e múltiplas saídas (MIMO, do inglês *multiple-input multiple-output*).

3.2.1 Modelo do Sistema com T_D para o Controle

O sistema de controle operando na taxa de amostragem original pode ser descrito pelo seguinte modelo em espaço de estados discreto:

$$\mathbf{x}_{k+1} = \mathbf{A}_D \mathbf{x}_k + \mathbf{B}_D \mathbf{u}_k^* + \mathbf{B}_w \mathbf{w}_k, \quad (3.1)$$

em que k representa o índice de tempo na taxa de amostragem original, $\mathbf{x}_k \in \mathbb{R}^n$ é o vetor de estados, $\mathbf{u}_k^* \in \mathbb{R}^m$ é o vetor de entradas de controle, e $\mathbf{w}_k \in \mathbb{R}^n$ é o vetor de ruído de processo que se assume ser gaussiano com média nula e matriz de covariância Σ_w , isto é, $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \Sigma_w)$. As matrizes $\mathbf{A}_D \in \mathbb{R}^{n \times n}$ e $\mathbf{B}_D \in \mathbb{R}^{n \times m}$ são obtidas por meio da discretização do modelo contínuo com período T_D .

A saída do sistema é dada por:

$$\mathbf{y}_k = \mathbf{C} \mathbf{x}_k + \mathbf{v}_k, \quad (3.2)$$

em que $\mathbf{C} \in \mathbb{R}^{p \times n}$ é a matriz de saída, $\mathbf{y}_k \in \mathbb{R}^p$ é o vetor de saídas medidas, e $\mathbf{v}_k \in \mathbb{R}^p$ representa o vetor de ruído de medição.

3.2.2 Modelo do Sistema com T_d para o Ataque

Para executar o ataque, o adversário utiliza um modelo do sistema discretizado com período de amostragem $T_d < T_D$. Definindo a razão de superamostragem como $r = T_D/T_d$, sendo r é um número inteiro positivo, o modelo superamostrado é descrito por:

$$\bar{\mathbf{x}}_{\ell+1} = \bar{\mathbf{A}}_d \bar{\mathbf{x}}_\ell + \bar{\mathbf{B}}_d \bar{\mathbf{u}}_\ell + \bar{\mathbf{B}}_w \bar{\mathbf{w}}_\ell, \quad (3.3)$$

em que o índice ℓ representa o instante de tempo na taxa de superamostragem, $\bar{\mathbf{x}}_\ell \in \mathbb{R}^n$ é o estado do sistema superamostrado, $\bar{\mathbf{u}}_\ell \in \mathbb{R}^m$ é o vetor de entradas superamostradas,

e as matrizes $\bar{\mathbf{A}}_d \in \mathbb{R}^{n \times n}$ e $\bar{\mathbf{B}}_d \in \mathbb{R}^{n \times m}$ são obtidas por meio da discretização do mesmo modelo contínuo com período T_d .

3.2.3 Condição de Indetectabilidade

O objetivo fundamental do atacante é garantir que, após r passos na taxa de superamostragem, o estado do sistema manipulado coincida exatamente com o estado que o sistema original teria alcançado nos instantes de amostragem. Essa condição de coincidência é expressa matematicamente como:

$$\bar{\mathbf{x}}_{\ell+r} = \mathbf{x}_{k+1}, \quad \text{com condição inicial} \quad \bar{\mathbf{x}}_\ell = \mathbf{x}_k. \quad (3.4)$$

Para desenvolver essa condição, vamos expandir a evolução do modelo superamostrado ao longo de r passos. Partindo do estado inicial $\bar{\mathbf{x}}_\ell = \mathbf{x}_k$ e aplicando recursivamente a Equação (3.3) desprezando o ruído de processo para simplificar a análise, obtemos:

$$\begin{aligned} \bar{\mathbf{x}}_{\ell+1} &= \bar{\mathbf{A}}_d \mathbf{x}_k + \bar{\mathbf{B}}_d \bar{\mathbf{u}}_\ell \\ \bar{\mathbf{x}}_{\ell+2} &= \bar{\mathbf{A}}_d \bar{\mathbf{x}}_{\ell+1} + \bar{\mathbf{B}}_d \bar{\mathbf{u}}_{\ell+1} \\ &= \bar{\mathbf{A}}_d^2 \mathbf{x}_k + \bar{\mathbf{A}}_d \bar{\mathbf{B}}_d \bar{\mathbf{u}}_\ell + \bar{\mathbf{B}}_d \bar{\mathbf{u}}_{\ell+1} \\ \bar{\mathbf{x}}_{\ell+3} &= \bar{\mathbf{A}}_d \bar{\mathbf{x}}_{\ell+2} + \bar{\mathbf{B}}_d \bar{\mathbf{u}}_{\ell+2} \\ &= \bar{\mathbf{A}}_d^3 \mathbf{x}_k + \bar{\mathbf{A}}_d^2 \bar{\mathbf{B}}_d \bar{\mathbf{u}}_\ell + \bar{\mathbf{A}}_d \bar{\mathbf{B}}_d \bar{\mathbf{u}}_{\ell+1} + \bar{\mathbf{B}}_d \bar{\mathbf{u}}_{\ell+2}. \end{aligned} \quad (3.5)$$

Generalizando esse padrão para r passos, chegamos a:

$$\bar{\mathbf{x}}_{\ell+r} = \bar{\mathbf{A}}_d^r \mathbf{x}_k + \sum_{s=1}^r \bar{\mathbf{A}}_d^{s-1} \bar{\mathbf{B}}_d \bar{\mathbf{u}}_{\ell+r-s}. \quad (3.6)$$

O primeiro termo, $\bar{\mathbf{A}}_d^r \mathbf{x}_k$, representa a evolução natural do estado inicial mediante a dinâmica do sistema superamostrado. O somatório representa a contribuição acumulada das ações de controle aplicadas durante r passos. Note que o índice s varia de 1 a r , e para cada valor de s , o termo $\bar{\mathbf{A}}_d^{s-1} \bar{\mathbf{B}}_d \bar{\mathbf{u}}_{\ell+r-s}$ representa o efeito da ação de controle $\bar{\mathbf{u}}_{\ell+r-s}$ propagado por $s-1$ passos de tempo até o instante final $\ell+r$.

Para que a condição de indetectabilidade seja satisfeita, igualamos a Equação (3.6) com o estado do sistema original no instante $k+1$:

$$\bar{\mathbf{A}}_d^r \mathbf{x}_k + \sum_{s=1}^r \bar{\mathbf{A}}_d^{s-1} \bar{\mathbf{B}}_d \bar{\mathbf{u}}_{\ell+r-s} = \mathbf{A}_D \mathbf{x}_k + \mathbf{B}_D \mathbf{u}_k^*. \quad (3.7)$$

Essa equação estabelece a restrição fundamental que as ações de controle do ataque devem satisfazer para permanecerem indetectáveis.

3.2.4 Otimização do Ataque

O problema de otimização do ataque é formulado em duas etapas sequenciais, cada uma com objetivos específicos e complementares. Essa abordagem garante tanto a factibilidade quanto a eficácia máxima do ataque.

3.2.4.1 Etapa 1: Busca de Solução Factível

A primeira etapa consiste em encontrar um conjunto de ações de controle $\{\bar{\mathbf{u}}_\ell, \bar{\mathbf{u}}_{\ell+1}, \dots, \bar{\mathbf{u}}_{\ell+r-1}\}$ que satisfaça a restrição de indetectabilidade expressa na Equação (3.7). Para sistemas MIMO, o vetor de ataque concatena todas as entradas de controle ao longo dos r passos superamostrados:

$$\mathbf{u}_a = \begin{bmatrix} \bar{\mathbf{u}}_\ell \\ \bar{\mathbf{u}}_{\ell+1} \\ \vdots \\ \bar{\mathbf{u}}_{\ell+r-1} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_{a,0} \\ \mathbf{u}_{a,1} \\ \vdots \\ \mathbf{u}_{a,r-1} \end{bmatrix} \in \mathbb{R}^{rm}, \quad (3.8)$$

em que $\mathbf{u}_{a,i} \in \mathbb{R}^m$ para cada $i = 0, \dots, r-1$.

Esse problema é formulado como um programa linear:

$$\begin{aligned} &\text{encontrar } \mathbf{u}_a \in \mathbb{R}^{rm} \\ &\text{sujeito a } \mathbf{C}_r \mathbf{u}_a = \mathbf{d}_k \\ &\quad \underline{\mathbf{u}} \leq \mathbf{u}_{a,i} \leq \bar{\mathbf{u}}, \quad i = 0, \dots, r-1, \end{aligned} \quad (3.9)$$

em que a matriz de alcançabilidade (ou controlabilidade) $\mathbf{C}_r \in \mathbb{R}^{n \times rm}$ é definida como:

$$\mathbf{C}_r = \begin{bmatrix} \bar{\mathbf{A}}_d^{r-1} \bar{\mathbf{B}}_d & \bar{\mathbf{A}}_d^{r-2} \bar{\mathbf{B}}_d & \dots & \bar{\mathbf{A}}_d \bar{\mathbf{B}}_d & \bar{\mathbf{B}}_d \end{bmatrix} \in \mathbb{R}^{n \times rm}. \quad (3.10)$$

Cada bloco $\bar{\mathbf{A}}_d^{r-1-i} \bar{\mathbf{B}}_d \in \mathbb{R}^{n \times m}$ corresponde à contribuição das m entradas no instante i , e o vetor alvo $\mathbf{d}_k \in \mathbb{R}^n$ é dado por:

$$\mathbf{d}_k = \mathbf{A}_D \mathbf{x}_k + \mathbf{B}_D \mathbf{u}_k^* - \bar{\mathbf{A}}_d^r \mathbf{x}_k, \quad (3.11)$$

em que $\mathbf{u}_k^* \in \mathbb{R}^m$ é o vetor de controle nominal (legítimo).

As restrições de saturação $\underline{\mathbf{u}}, \bar{\mathbf{u}} \in \mathbb{R}^m$ são os vetores de limites inferior e superior para cada entrada, aplicados componente a componente:

$$[\underline{\mathbf{u}}]_j \leq [\mathbf{u}_{a,i}]_j \leq [\bar{\mathbf{u}}]_j, \quad \forall i \in \{0, \dots, r-1\}, \forall j \in \{1, \dots, m\}. \quad (3.12)$$

Análise dimensional: O sistema possui n equações de igualdade (coincidência) e rm variáveis de decisão, resultando em $rm - n$ graus de liberdade para otimização. Adicionalmente, há $2rm$ restrições de desigualdade (saturação).

3.2.4.2 Etapa 2: Maximização da Energia

Uma vez obtida uma solução factível $\mathbf{u}_{\text{ini}}$ na primeira etapa, procede-se à maximização da energia da ação de controle. Para sistemas MIMO, a energia total considera a contribuição de todas as m entradas:

$$\begin{aligned} \max_{\mathbf{u}_a} \quad & E(\mathbf{u}_a) = \mathbf{u}_a^\top \mathbf{u}_a = \sum_{i=0}^{r-1} \|\mathbf{u}_{a,i}\|^2 = \sum_{i=0}^{r-1} \sum_{j=1}^m [u_{a,i}]_j^2 \\ \text{sujeito a} \quad & \mathbf{C}_r \mathbf{u}_a = \mathbf{d}_k \\ & \underline{\mathbf{u}} \leq \mathbf{u}_{a,i} \leq \bar{\mathbf{u}}, \quad i = 0, \dots, r-1. \end{aligned} \quad (3.13)$$

Note que maximizar $\mathbf{u}_a^\top \mathbf{u}_a$ é equivalente a minimizar $-\mathbf{u}_a^\top \mathbf{u}_a$. Essa reformulação permite o uso de algoritmos de minimização padrão. O problema é resolvido utilizando métodos de programação quadrática sequencial (SQP), conforme descrito por Boggs e Tolle (1995) e Gill, Murray e Saunders (2005).

O algoritmo SQP aproxima o problema não linear original por uma sequência de subproblemas quadráticos, cada um resolvido usando métodos de região de confiança ou busca linear. A solução factível $\mathbf{u}_{\text{ini}}$ obtida na primeira etapa serve como ponto inicial para o algoritmo, garantindo que a busca comece em uma região viável do espaço de soluções.

3.2.4.3 Implementação Computacional

É importante destacar que o problema de otimização deve ser resolvido a cada instante de amostragem na taxa T_d utilizada pelo atacante. Isso significa que a cada novo passo do sistema superamostrado, um novo conjunto de ações de controle deve ser computado, garantindo que a condição de indetectabilidade seja continuamente satisfeita enquanto se maximiza o impacto do ataque.

A complexidade computacional do algoritmo completo é ligada à solução do problema quadrático na segunda etapa, resultando em complexidade $\mathcal{O}((rm)^3)$ por iteração (NOCEDAL; WRIGHT, 2006).

3.2.5 Quantificação do Esforço do Ataque

Para avaliar quantitativamente o impacto do ataque de superamostragem sobre o sistema, é fundamental estabelecer métricas que capturem o aumento no consumo energético e o desgaste adicional imposto aos atuadores considerando todas as m entradas de controle.

3.2.5.1 Métrica de Energia da Raiz Quadrada da Média (RMS, do inglês *root mean square*)

A métrica principal utilizada para quantificar o esforço do ataque nesses sistemas é a energia da raiz quadrada média (RMS) da ação de controle, definida como:

$$E_{\text{RMS}} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\mathbf{u}_a(i)\|^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^m [u_a(i)]_j^2}, \quad (3.14)$$

em que N é o número total de amostras consideradas no período de avaliação, $\mathbf{u}_a(i) \in \mathbb{R}^m$ representa o vetor de ações de controle no instante i na taxa de superamostragem, e $[u_a(i)]_j$ denota a j -ésima componente deste vetor.

Essa métrica é utilizada na análise de sistemas de controle por fornecer uma medida normalizada do esforço de controle agregado em todas as entradas (SKOGESTAD; POSTLETHWAITE, 2005). O valor RMS captura tanto a magnitude quanto a persistência das ações de controle em cada atuador, sendo diretamente proporcional à energia elétrica total consumida em atuadores eletromecânicos.

3.2.5.2 Taxa de Incremento Energético

Para comparar o consumo energético entre o sistema normal e o sistema sob ataque, define-se a taxa de incremento energético percentual:

$$\Delta E(\%) = \frac{E_{\text{RMS}}^{\text{ataque}} - E_{\text{RMS}}^{\text{normal}}}{E_{\text{RMS}}^{\text{normal}}} \times 100. \quad (3.15)$$

Essa métrica fornece uma medida relativa do impacto do ataque, independente da escala absoluta do sistema e do número de atuadores.

3.2.5.3 Índice de Desgaste do Atuador

Além do consumo energético, o ataque de superamostragem impõe ciclos de operação mais frequentes aos atuadores, acelerando seu desgaste mecânico. Para esses sistemas, o índice de desgaste considera a variação total em todos os m atuadores:

$$TV = \sum_{i=1}^{N-1} \|\mathbf{u}_a(i+1) - \mathbf{u}_a(i)\|_1 = \sum_{i=1}^{N-1} \sum_{j=1}^m |[u_a(i+1)]_j - [u_a(i)]_j|. \quad (3.16)$$

Esse índice captura a quantidade de mudanças nas ações de controle de todos os atuadores, sendo proporcional ao desgaste mecânico agregado em válvulas, motores e outros atuadores sujeitos a fadiga por ciclagem (QIN; BADGWELL, 2003; HARRIS; SEPPALA; DESBOROUGH, 1999). O aumento na variação total sob ataque pode levar à falha prematura de componentes, comprometendo a confiabilidade e disponibilidade do sistema de controle.

3.2.6 Detecção do Ataque de Superamostragem

Uma das estratégias utilizadas para identificar o ataque de superamostragem consiste na utilização do filtro de *Kalman* operando com uma taxa de amostragem superior à do sistema original. Essa abordagem permite detectar diferenças na dinâmica do sistema que permanecem ocultas quando o monitoramento ocorre apenas nos instantes de amostragem nominais.

3.2.6.1 Configuração do Filtro de *Kalman* Superamostrado

O filtro de *Kalman* é configurado para operar com período $T_{KF} = T_D/N$, em que $N \in \mathbb{N}^*$ é um inteiro maior que 1, permitindo observar o comportamento do sistema entre os instantes de amostragem originais. Quando o filtro opera na mesma taxa de amostragem do sistema ($T_{KF} = T_D$), o ataque permanece indetectável, pois as manipulações introduzidas ocorrem justamente entre esses instantes de observação. Ao aumentar a frequência de operação do filtro, cria-se a capacidade de capturar estados intermediários do sistema que revelam as diferenças introduzidas pelo ataque.

O modelo linear discreto adotado para o filtro, considerando múltiplas entradas e múltiplas saídas (MIMO), é dado por:

$$\begin{cases} \mathbf{x}_{j+1} = \bar{\bar{\mathbf{A}}}\mathbf{x}_j + \bar{\bar{\mathbf{B}}}\mathbf{u}_j + \bar{\bar{\mathbf{B}}}_w\mathbf{w}_j \\ \mathbf{y}_{j+1} = \bar{\bar{\mathbf{C}}}\mathbf{x}_{j+1} + \mathbf{v}_{j+1}, \end{cases} \quad (3.17)$$

em que o índice j representa os instantes de tempo na taxa de superamostragem do filtro, $\mathbf{u}_j \in \mathbb{R}^m$ é o vetor de entradas de controle, as matrizes $\bar{\bar{\mathbf{A}}} \in \mathbb{R}^{n \times n}$, $\bar{\bar{\mathbf{B}}} \in \mathbb{R}^{n \times m}$ e $\bar{\bar{\mathbf{C}}} \in \mathbb{R}^{p \times n}$ são obtidas por meio da discretização do modelo contínuo com período T_{KF} . O vetor de ruído de processo $\mathbf{w}_j \in \mathbb{R}^n$ possui matriz de covariância $\bar{\bar{\mathbf{Q}}} = \mathbb{E}[\mathbf{w}_j\mathbf{w}_j^T]$, e o vetor de ruído de medição $\mathbf{v}_{j+1} \in \mathbb{R}^p$ possui matriz de covariância $\bar{\bar{\mathbf{R}}} = \mathbb{E}[\mathbf{v}_{j+1}\mathbf{v}_{j+1}^T]$.

3.2.6.2 Implementação do Filtro de *Kalman*

A implementação do filtro de *Kalman* segue o ciclo clássico de predição e correção (KALMAN, 1960; MAYBECK, 1982), adaptado aqui para operar na taxa de superamostragem T_{KF} . A cada instante j , o filtro primeiro propaga o estado estimado com base no modelo dinâmico e, em seguida, corrige essa estimativa utilizando a medição disponível, atualizando também as respectivas matrizes de covariância do erro.

Etapa de Propagação

A etapa de propagação do filtro de *Kalman* prediz o estado do sistema e sua incerteza para o próximo instante, baseando-se no modelo dinâmico. A predição do estado

para o tempo t_{j+1} é dada por:

$$\hat{\mathbf{x}}_{j+1}^- = \bar{\bar{\mathbf{A}}} \hat{\mathbf{x}}_j^+ + \bar{\bar{\mathbf{B}}} \mathbf{u}_j, \quad (3.18)$$

em que $\hat{\mathbf{x}}_j^+$ é a estimativa a posteriori no instante j (após correção com a medição) e $\hat{\mathbf{x}}_{j+1}^-$ é a estimativa a priori no instante $j+1$ (antes da correção). A matriz de covariância do erro de estimação é propagada segundo:

$$\mathbf{P}_{j+1}^- = \bar{\bar{\mathbf{A}}} \mathbf{P}_j^+ \bar{\bar{\mathbf{A}}}^\top + \bar{\bar{\mathbf{B}}} \bar{\bar{\mathbf{Q}}} \bar{\bar{\mathbf{B}}}^\top, \quad (3.19)$$

quantificando como a incerteza se propaga através da dinâmica do sistema, incorporando a contribuição do ruído de processo. A derivação completa dessa equação é apresentada por Crassidis e Junkins (2004).

Etapa de Correção

A etapa de correção introduz informações das medições para refinar a estimativa do estado. O ganho de *Kalman*, que representa o peso ótimo para combinar predição e medição, é calculado como:

$$\mathbf{K}_{j+1} = \mathbf{P}_{j+1}^- \bar{\bar{\mathbf{C}}}^\top \left[\bar{\bar{\mathbf{C}}} \mathbf{P}_{j+1}^- \bar{\bar{\mathbf{C}}}^\top + \bar{\bar{\mathbf{R}}} \right]^{-1}, \quad (3.20)$$

minimizando o traço da matriz de covariância do erro a posteriori. A estimativa do estado é então corrigida por:

$$\hat{\mathbf{x}}_{j+1}^+ = \hat{\mathbf{x}}_{j+1}^- + \mathbf{K}_{j+1} \left[\mathbf{y}_{j+1} - \bar{\bar{\mathbf{C}}} \hat{\mathbf{x}}_{j+1}^- \right], \quad (3.21)$$

sendo o termo $\mathbf{y}_{j+1} - \bar{\bar{\mathbf{C}}} \hat{\mathbf{x}}_{j+1}^-$ a inovação, que quantifica a diferença entre medição observada e predição. A matriz de covariância do erro de estimação é atualizada para representar a redução na incerteza:

$$\mathbf{P}_{j+1}^+ = \left(\mathbf{I} - \mathbf{K}_{j+1} \bar{\bar{\mathbf{C}}} \right) \mathbf{P}_{j+1}^-, \quad (3.22)$$

conhecida como forma de *Joseph*, que garante que a matriz permaneça positiva semi-definida mesmo na presença de erros numéricos (BUCY; JOSEPH, 2005).

Disponibilidade dos Dados Superamostrados

Para que o filtro de *Kalman* possa realizar a etapa de correção nos instantes intermediários j , é necessário que as medições $\mathbf{y}_{j+1}$ estejam disponíveis na taxa T_{KF} . Isso significa que o sistema de monitoramento deve ter acesso aos dados do processo superamostrado, ou seja, deve ser capaz de adquirir medições dos sensores em uma frequência superior à do controlador original. Na prática, isso pode ser viabilizado por meio de um sistema de aquisição de dados operando em paralelo ao controlador, sem que este precise ser modificado.

3.2.7 Detecção via Análise de Erro

Para identificar a presença do ataque de superamostragem, emprega-se um detector de anomalias baseado na análise estatística do erro de estimação do filtro de *Kalman*, utilizando teste de hipóteses sequencial (BRUMBACK; SRINATH, 1987). O resíduo ou erro de estimação no instante j é definido como:

$$\mathbf{z}_j = \mathbf{y}_j - \bar{\mathbf{C}}\hat{\mathbf{x}}_j^+, \quad (3.23)$$

representando a diferença entre a medição observada e a estimativa do filtro. Sob condições normais de operação (hipótese nula $\mathbf{H}_0$), a sequência de resíduos $\mathbf{z}_j$ constitui um processo branco gaussiano:

$$\mathbf{z}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_j), \quad \mathbb{E}[\mathbf{z}_i \mathbf{z}_k^\top] = \mathbf{S}_j \delta_{ik}, \quad \text{sob } \mathbf{H}_0, \quad (3.24)$$

sendo δ_{ik} o delta de *Kronecker* e a matriz de covariância do resíduo dada por:

$$\mathbf{S}_j = \bar{\mathbf{C}}\mathbf{P}_j^-\bar{\mathbf{C}}^\top + \bar{\mathbf{R}}. \quad (3.25)$$

O índice de detecção baseado na estatística qui-quadrado é calculado como:

$$g(j) = \mathbf{z}_j^\top \mathbf{S}_j^{-1} \mathbf{z}_j, \quad (3.26)$$

seguindo uma distribuição qui-quadrado com p graus de liberdade, em que p é a dimensão do vetor de medições:

$$g(j) \sim \chi_p^2, \quad \text{sob } \mathbf{H}_0. \quad (3.27)$$

A detecção do ataque ocorre quando $g(j) > \Psi$, sendo Ψ o limiar escolhido para garantir uma probabilidade de falso alarme especificada. O limiar pode ser determinado por abordagem heurística baseada na separação de envelopes de dados, adotando-se a média geométrica entre valores extremos observados para equilibrar a margem de decisão em escala logarítmica (ISERMANN, 2005):

$$\Psi_{\chi^2} = \sqrt{g_{\max}^{\text{normal}} \cdot g_{\min}^{\text{ataque}}}, \quad (3.28)$$

garantindo separação balanceada entre regimes normal e sob ataque, minimizando falsos positivos e negativos. Alternativamente, o limiar pode ser determinado mediante o critério de *Neyman-Pearson*, fixando-se a probabilidade de falso alarme α (KAY, 1993):

$$\Psi_{\chi^2} = \chi_{p,1-\alpha}^2, \quad (3.29)$$

em que $\chi_{p,1-\alpha}^2$ denota o quartil $(1 - \alpha)$ da distribuição qui-quadrado com p graus de liberdade.

3.2.8 Detecção via Análise da ACF do Erro

Uma abordagem complementar para detecção do ataque baseia-se na análise da ACF do resíduo, explorando o fato de que, sob condições normais, o resíduo deve comportar-se como processo de ruído branco (MEHRA, 1970). A função de autocorrelação amostral do resíduo para um atraso τ é calculada como:

$$\rho_z(\tau) = \frac{\sum_{j=1}^{N-\tau} (\mathbf{z}_j - \bar{\mathbf{z}})^\top (\mathbf{z}_{j+\tau} - \bar{\mathbf{z}})}{\sum_{j=1}^N (\mathbf{z}_j - \bar{\mathbf{z}})^\top (\mathbf{z}_j - \bar{\mathbf{z}})}, \quad (3.30)$$

em que $\bar{\mathbf{z}}$ é a média amostral do resíduo e N é o tamanho da janela de análise. Sob a hipótese de que o resíduo é um processo branco, os valores da autocorrelação para $\tau \neq 0$ devem ser estatisticamente nulos. O teste de *Ljung-Box* fornece uma estatística agregada para testar essa hipótese (LJUNG; BOX, 1978):

$$Q = N(N+2) \sum_{\tau=1}^h \frac{\rho_z^2(\tau)}{N-\tau}, \quad (3.31)$$

sendo h o número máximo de atrasos considerados sob a hipótese nula $Q \sim \chi_h^2$. Para detecção em tempo real, utiliza-se janela deslizante de tamanho W sobre o resíduo. A cada instante j , calcula-se:

$$\rho_{\max}(j) = \max_{\tau \in \{1, \dots, \tau_{\max}\}} |\rho_z(\tau)|_{[j-W+1:j]}, \quad (3.32)$$

disparando-se o alarme quando $\rho_{\max}(j) > \Psi_{\text{ACF}}$.

Para um nível de confiança de $(1 - \alpha)$, o limiar pode ser estabelecido como:

$$\Psi_{\text{ACF}} = \frac{z_{1-\alpha/2}}{\sqrt{W}}, \quad (3.33)$$

sendo $z_{1-\alpha/2}$ o quantil da distribuição normal padrão. Para $\alpha = 0,001$ (confiança de 99,9%), tem-se $z_{0,999} \approx 3,291$, resultando em $\Psi_{\text{ACF}} \approx 3,291/\sqrt{W}$ (BROCKWELL; DAVIS, 2002).

3.2.9 Análise Comparativa dos Métodos de Detecção

Os dois métodos de detecção apresentam características complementares que justificam sua aplicação conjunta:

- **Detecção via χ^2 :** Oferece resposta rápida a mudanças rápidas no comportamento do sistema, sendo particularmente eficiente na detecção inicial do ataque. A sensibilidade pode ser ajustada mediante o limiar Ψ .
- **Detecção via ACF:** Mais robusta a perturbações transitórias, identifica padrões persistentes de correlação temporal introduzidos pelo ataque. Requer maior volume de dados para detecção confiável.

Essa abordagem dual tem sido validada em aplicações práticas de detecção de anomalias em sistemas de controle (ISERMANN, 2005).

3.3 Extensão para Sistemas Não Lineares

A formulação apresentada na Seção 3.2 considera sistemas lineares invariantes no tempo, para os quais a condição de furtividade resulta em uma restrição linear convexa. Entretanto, sistemas de controle industrial frequentemente apresentam dinâmicas não lineares, seja por características intrínsecas do processo (saturações, zonas mortas, dinâmicas de fluidos) ou por condições de operação variáveis.

Nesta seção, apresentam-se duas metodologias para estender o ataque de superamostragem a sistemas não lineares: a modelagem *fuzzy Takagi-Sugeno*, que interpola subsistemas lineares locais por meio de funções de pertinência, e a teoria dos Operadores de *Koopman*, que eleva o sistema a um espaço de observáveis onde a dinâmica é globalmente linear. As abordagens apresentadas transformam o problema não linear original em um problema linear (ou *quasi*-linear) equivalente, mantendo a estrutura de programação quadrática do problema de otimização do ataque.

3.3.1 Formulação do Ataque via Modelagem *Takagi-Sugeno*

A modelagem *fuzzy Takagi-Sugeno* (T-S) permite representar sistemas não lineares como uma interpolação de modelos lineares locais, ponderados por funções de pertinência que dependem das variáveis premissa do sistema (TAKAGI; SUGENO, 1985).

3.3.1.1 Sistema Original Não Linear

Considera-se um sistema não linear afim no controle, descrito por:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) + \mathbf{G}(\mathbf{x}(t))\mathbf{u}(t) + \mathbf{B}_w\mathbf{w}(t), \quad (3.34)$$

em que $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ e $\mathbf{G} : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ são funções não lineares suaves.

Representação T-S via Não Linearidade Setorial

A técnica de não linearidade setorial permite representar exatamente o sistema descrito pela Equação (3.34) como uma combinação convexa de subsistemas lineares, desde que as não linearidades sejam limitadas em uma região compacta $\mathcal{D} \subset \mathbb{R}^n$ (TANAKA; WANG, 2004).

Para isso, reescreve-se o sistema descrito na Equação (3.34) na forma *quasi*-LPV (do inglês, *linear parameter-varying*):

$$\dot{\mathbf{x}}(t) = \mathbf{A}(\mathbf{z}(t))\mathbf{x}(t) + \mathbf{B}(\mathbf{z}(t))\mathbf{u}(t) + \mathbf{B}_w\mathbf{w}(t), \quad (3.35)$$

em que $\mathbf{z}(t) = [z_1(t), \dots, z_p(t)]^\top$ contém os termos não lineares.

Para cada não linearidade $z_i(\mathbf{x})$, definem-se os limites:

$$\underline{z}_i = \min_{\mathbf{x} \in \mathcal{D}} z_i(\mathbf{x}), \quad \bar{z}_i = \max_{\mathbf{x} \in \mathcal{D}} z_i(\mathbf{x}). \quad (3.36)$$

As funções de pertinência são:

$$M_i^1(z_i) = \frac{\bar{z}_i - z_i}{\bar{z}_i - \underline{z}_i}, \quad M_i^2(z_i) = \frac{z_i - \underline{z}_i}{\bar{z}_i - \underline{z}_i}. \quad (3.37)$$

Com p não linearidades, obtém-se $r_{\text{regras}} = 2^p$ regras *fuzzy*:

$$\dot{\mathbf{x}}(t) = \sum_{i=1}^{r_{\text{regras}}} h_i(\mathbf{z}(t)) [\mathbf{A}_i \mathbf{x}(t) + \mathbf{B}_i \mathbf{u}(t)] + \mathbf{B}_w \mathbf{w}(t), \quad (3.38)$$

em que $\mathbf{A}_i \in \mathbb{R}^{n \times n}$ e $\mathbf{B}_i \in \mathbb{R}^{n \times m}$ são as matrizes do i -ésimo subsistema linear, e as funções de pertinência normalizadas satisfazem:

$$\sum_{i=1}^{r_{\text{regras}}} h_i(\mathbf{z}) = 1, \quad h_i(\mathbf{z}) \geq 0, \quad \forall \mathbf{z} \in \mathcal{D}. \quad (3.39)$$

3.3.1.2 Discretização do Modelo T-S

Para aplicação em sistemas de controle digital, o modelo T-S contínuo expressa na Equação (3.38) deve ser discretizado (método ZOH, do inglês *zero-order hold*). O controlador legítimo opera com período de amostragem T_D , resultando no modelo discreto:

$$\mathbf{x}_{k+1} = \sum_{i=1}^{r_{\text{regras}}} h_i(\mathbf{z}_k) [\mathbf{A}_{D,i} \mathbf{x}_k + \mathbf{B}_{D,i} \mathbf{u}_k^*], \quad (3.40)$$

em que

$$\mathbf{A}_{D,i} = e^{\mathbf{A}_i T_D}, \quad \mathbf{B}_{D,i} = \left(\int_0^{T_D} e^{\mathbf{A}_i \tau} d\tau \right) \mathbf{B}_i, \quad (3.41)$$

são as matrizes discretizadas do i -ésimo subsistema local, com $\mathbf{A}_{D,i} \in \mathbb{R}^{n \times n}$ e $\mathbf{B}_{D,i} \in \mathbb{R}^{n \times m}$.

Observação: Na discretização via ZOH, assume-se que as funções de pertinência $h_i(\mathbf{z}(t))$ permanecem constantes ao longo do intervalo de amostragem $[kT_D, (k+1)T_D]$, ou seja, $h_i(\mathbf{z}(t)) \approx h_i(\mathbf{z}_k)$ para $t \in [kT_D, (k+1)T_D]$.

De forma compacta, definem-se as matrizes ponderadas:

$$\mathbf{A}_D(\mathbf{z}_k) = \sum_{i=1}^{r_{\text{regras}}} h_i(\mathbf{z}_k) \mathbf{A}_{D,i} \in \mathbb{R}^{n \times n}, \quad (3.42)$$

$$\mathbf{B}_D(\mathbf{z}_k) = \sum_{i=1}^{r_{\text{regras}}} h_i(\mathbf{z}_k) \mathbf{B}_{D,i} \in \mathbb{R}^{n \times m}, \quad (3.43)$$

resultando no modelo discreto dependente de parâmetros:

$$\mathbf{x}_{k+1} = \mathbf{A}_D(\mathbf{z}_k) \mathbf{x}_k + \mathbf{B}_D(\mathbf{z}_k) \mathbf{u}_k^*. \quad (3.44)$$

Para sistemas *fuzzy* T-S, o controlador de Compensação Distribuída Paralela (PDC, do inglês *parallel distributed compensation*) é tipicamente empregado por realimentação de estado:

$$\mathbf{u}_k^* = - \sum_{i=1}^{r_{\text{regras}}} h_i(\mathbf{z}_k) \mathbf{K}_i \mathbf{x}_k = -\mathbf{K}(h_k) \mathbf{x}_k, \quad (3.45)$$

em que $\mathbf{K}_i \in \mathbb{R}^{m \times n}$ é o ganho de controle associado ao i -ésimo subsistema.

3.3.1.3 Modelo do Atacante T-S

O atacante opera com período $T_d < T_D$:

$$T_d = \frac{T_D}{r}, \quad r \in \mathbb{N}, \quad r > 1, \quad (3.46)$$

sendo r é o fator de superamostragem.

Discretizando o modelo T-S com período T_d :

$$\bar{\mathbf{x}}_{\ell+1} = \sum_{i=1}^{r_{\text{regras}}} h_i(\bar{\mathbf{z}}_\ell) [\bar{\mathbf{A}}_{d,i} \bar{\mathbf{x}}_\ell + \bar{\mathbf{B}}_{d,i} \bar{\mathbf{u}}_\ell] + \bar{\mathbf{B}}_w \bar{\mathbf{w}}_\ell, \quad (3.47)$$

com

$$\bar{\mathbf{A}}_{d,i} = e^{\mathbf{A}_i T_d} \in \mathbb{R}^{n \times n}, \quad \bar{\mathbf{B}}_{d,i} = \left(\int_0^{T_d} e^{\mathbf{A}_i \tau} d\tau \right) \mathbf{B}_i \in \mathbb{R}^{n \times m}. \quad (3.48)$$

Definindo as matrizes agregadas superamostradas:

$$\bar{\mathbf{A}}_d(h_\ell) = \sum_{i=1}^{r_{\text{regras}}} h_i(\bar{\mathbf{z}}_\ell) \bar{\mathbf{A}}_{d,i}, \quad \bar{\mathbf{B}}_d(h_\ell) = \sum_{i=1}^{r_{\text{regras}}} h_i(\bar{\mathbf{z}}_\ell) \bar{\mathbf{B}}_{d,i}. \quad (3.49)$$

Hipótese 1 (Funções de Pertinências Fixas): Para T_d suficientemente pequeno:

$$h_i(\bar{\mathbf{z}}_{\ell+j}) \approx h_i(\mathbf{z}_k) = h_{i,k}, \quad \forall j \in \{0, \dots, r-1\}. \quad (3.50)$$

Sob essa hipótese, o vetor de ataque concatena todas as m entradas de controle ao longo dos r passos superamostrados:

$$\mathbf{u}_a = \begin{bmatrix} \bar{\mathbf{u}}_\ell \\ \bar{\mathbf{u}}_{\ell+1} \\ \vdots \\ \bar{\mathbf{u}}_{\ell+r-1} \end{bmatrix} \in \mathbb{R}^{rm}, \quad (3.51)$$

em que $\bar{\mathbf{u}}_{\ell+j} \in \mathbb{R}^m$ para cada $j = 0, \dots, r-1$.

A exigência de que a evolução do sistema superamostrado coincida com a evolução nominal ao final do intervalo de amostragem impõe a seguinte restrição:

$$[\bar{\mathbf{A}}_d(h_k)]^r \mathbf{x}_k + \sum_{j=0}^{r-1} [\bar{\mathbf{A}}_d(h_k)]^{r-1-j} \bar{\mathbf{B}}_d(h_k) \bar{\mathbf{u}}_{\ell+j} = \mathbf{A}_D(h_k) \mathbf{x}_k + \mathbf{B}_D(h_k) \mathbf{u}_k^*. \quad (3.52)$$

Reescrevendo a Equação (3.52) na forma matricial, obtém-se a condição de indelectabilidade:

$$\mathbf{C}_r(h_k)\mathbf{u}_a = \mathbf{d}_k(h_k), \quad (3.53)$$

sendo a matriz de controlabilidade superamostrada $\mathbf{C}_r(h_k) \in \mathbb{R}^{n \times rm}$ definida como:

$$\mathbf{C}_r(h_k) = \begin{bmatrix} [\bar{\mathbf{A}}_d(h_k)]^{r-1} \bar{\mathbf{B}}_d(h_k) & [\bar{\mathbf{A}}_d(h_k)]^{r-2} \bar{\mathbf{B}}_d(h_k) & \cdots & \bar{\mathbf{A}}_d(h_k) \bar{\mathbf{B}}_d(h_k) & \bar{\mathbf{B}}_d(h_k) \end{bmatrix}, \quad (3.54)$$

em que cada bloco $[\bar{\mathbf{A}}_d(h_k)]^{r-1-j} \bar{\mathbf{B}}_d(h_k) \in \mathbb{R}^{n \times m}$ corresponde à contribuição das m entradas no instante j , e o vetor alvo $\mathbf{d}_k(h_k) \in \mathbb{R}^n$ é dado por:

$$\mathbf{d}_k(h_k) = \left[\mathbf{A}_D(h_k) - [\bar{\mathbf{A}}_d(h_k)]^r \right] \mathbf{x}_k + \mathbf{B}_D(h_k) \mathbf{u}_k^*. \quad (3.55)$$

Importância da Hipótese 1: A Hipótese 1 é fundamental para viabilizar a solução do problema de ataque, pois permite transformar um problema de otimização com restrições variantes no tempo em um problema com restrições fixas, linearizando a condição de coincidência.

Justificativa matemática: Sem a Hipótese 1, a evolução do sistema superamostrado seria:

$$\bar{\mathbf{x}}_{\ell+r} = \prod_{j=0}^{r-1} \bar{\mathbf{A}}_d(h_{\ell+r-1-j}) \mathbf{x}_k + \sum_{j=0}^{r-1} \left[\prod_{s=j+1}^{r-1} \bar{\mathbf{A}}_d(h_{\ell+s}) \right] \bar{\mathbf{B}}_d(h_{\ell+j}) \bar{\mathbf{u}}_{\ell+j}, \quad (3.56)$$

em que cada $h_{\ell+j}$ depende de $\bar{\mathbf{x}}_{\ell+j}$, que por sua vez depende de todas as entradas de ataque anteriores. Isso resulta em um problema de otimização não linear acoplado com restrições implícitas. Assim, assumindo a condição expressa na Equação (3.50), obtém-se o lado esquerdo da Equação (3.52). Isso permite escrever a condição de coincidência expressa na Equação (3.53) como um sistema linear.

A hipótese é válida quando:

- **Superamostragem alta:** $T_d = T_D/r \ll 1$. Para T_d pequeno, o estado varia pouco entre amostras; consecutivas, logo $\bar{\mathbf{z}}_{\ell+j} \approx \mathbf{z}_k$.
- **Não linearidades suaves:** As funções de pertinência $h_i(\mathbf{z}(\mathbf{x}))$ devem variar lentamente em relação ao período de superamostragem T_d , de modo que a hipótese expressa na Equação (3.50) seja satisfeita. Não linearidades com variação rápida no estado, mesmo que contínuas, podem invalidar essa aproximação.

3.3.1.4 Problema de Otimização do Ataque

A formulação do problema de otimização para sistemas não lineares T-S MIMO mantém a estrutura do problema expressa na Equação (3.13), substituindo-se:

- $\mathbf{C}_r$ por $\mathbf{C}_r(h_k)$;

- $\mathbf{d}_k$ por $\mathbf{d}_k(h_k)$.

O problema de otimização formado é não convexo. De acordo com o Corolário 32.3.4 de Rockafellar (1997), se uma função convexa é maximizada sobre um conjunto convexo poliédrico limitado (que não contém linhas completas) e a função é limitada superiormente nesse conjunto, então o supremo é atingido em um dos pontos extremos do conjunto. Consequentemente, garante-se que a solução ótima global $\mathbf{u}_a^*$ reside em um dos vértices do politopo, o que permite restringir o espaço de busca a um conjunto finito de pontos.

3.3.1.5 Métodos de Solução

Para resolver o problema de otimização do ataque em sistemas não lineares T-S MIMO, quatro métodos podem ser empregados, sendo o método 4 utilizado neste trabalho. Os métodos são:

1. **Formulação via Enumeração e Projeção de Vértices:** Explora a propriedade de que o máximo global de uma função convexa é atingido em um de seus vértices (ROCKAFELLAR, 1997), enumerando todos os 2^m vértices do hipercubo de restrições e calculando suas projeções no subespaço linear definido pelas restrições de igualdade;
2. **Formulação via Relaxação Semidefinida (SDP):** Convexifica o problema não-convexo mediante relaxação semidefinida, elevando a variável $\mathbf{u}_a$ ao espaço de matrizes simétricas e resolvendo um problema de programação semidefinida (BOYD; VANDENBERGHE, 2004);
3. **Formulação via Espaço Nulo:** Parametriza o conjunto de soluções factíveis usando decomposição em espaço nulo, eliminando as restrições de igualdade e reduzindo o problema a otimização em espaço de dimensão reduzida (NOCEDAL; WRIGHT, 2006);
4. **Formulação via Otimização Não Linear com Multi-Start Guiado por GRASP:** Combina a parametrização do espaço nulo com otimização não linear, utilizando múltiplos pontos iniciais gerados via procedimentos de busca adaptativa aleatorizada gulosa (GRASP, do inglês *Greedy Randomized Adaptive Search Procedures*) (FEO; RESENDE, 1995).

Formulação via Otimização Não Linear com Multi-Start Guiado por GRASP: O método é executado em três etapas sequenciais. Primeiramente, calcula-se a solução particular de mínima norma que satisfaz a restrição linear:

$$\mathbf{u}_a^{(0)} = \mathbf{C}_r^\top (\mathbf{C}_r \mathbf{C}_r^\top)^{-1} \mathbf{d}_k, \quad (3.57)$$

fornecendo uma solução viável que minimiza a energia do ataque. A solução geral é então parametrizada utilizando a base do espaço nulo $\mathbf{N}_r \in \mathbb{R}^{rm \times (rm-n)}$, que representa o núcleo de $\mathcal{C}_r$:

$$\mathbf{u}_a = \mathbf{u}_a^{(0)} + \mathbf{N}_r \boldsymbol{\eta}, \quad (3.58)$$

em que $\boldsymbol{\eta} \in \mathbb{R}^{rm-n}$ é o vetor de parâmetros livres que permite explorar o espaço de soluções factíveis. O problema de maximização torna-se:

$$\begin{aligned} \max_{\boldsymbol{\eta}} \quad & J(\boldsymbol{\eta}) = \|\mathbf{u}_a^{(0)} + \mathbf{N}_r \boldsymbol{\eta}\|^2, \\ \text{sujeito a} \quad & \underline{\mathbf{u}} \leq [\mathbf{u}_a^{(0)} + \mathbf{N}_r \boldsymbol{\eta}]_{(j-1)m+1:jm} \leq \bar{\mathbf{u}}, \quad j = 1, \dots, r. \end{aligned} \quad (3.59)$$

O algoritmo realiza $N_{\text{tentativas}}$ iterações, gerando em cada iteração um ponto inicial $\boldsymbol{\eta}_0^{(i)}$ via GRASP e refinando-o mediante otimização não linear. A metodologia GRASP é composta por duas fases complementares. Na Fase 1 (Construção Adaptativa Gulosa), o vetor $\boldsymbol{\eta}_0^{(i)}$ é construído componente a componente ($s = 1, \dots, rm - n$). Para cada componente, define-se um conjunto finito de candidatos $\mathcal{D}_s$ mediante discretização do intervalo viável. Calcula-se então uma função de mérito $g_s(\nu)$ que mensura o impacto de cada candidato $\nu \in \mathcal{D}_s$ na energia final:

$$g_s(\nu) = \|\mathbf{u}_a^{(s-1)} + \nu \mathbf{n}_s\|^2, \quad (3.60)$$

sendo $\mathbf{u}_a^{(s-1)}$ o vetor de ataque construído com as componentes já fixadas e $\mathbf{n}_s$ a s -ésima coluna de $\mathbf{N}_r$. Os melhores candidatos são filtrados mediante a lista restrita de candidatos (RCL, do inglês *restricted candidate list*):

$$\text{RCL} = \{\nu \in \mathcal{D}_s \mid g_s(\nu) \geq g_{\max} - \alpha(g_{\max} - g_{\min})\}, \quad (3.61)$$

em que $g_{\max}$ e $g_{\min}$ são os valores máximo e mínimo da função de mérito, e o parâmetro $\alpha \in [0,1]$ controla o nível de aleatoriedade: $\alpha = 0$ corresponde a construção puramente gulosa (determinística), enquanto $\alpha = 1$ equivale a escolha puramente aleatória. Seleciona-se então aleatoriamente um valor da RCL para a componente η_s . Na Fase 2 (Busca Local), o vetor construído $\boldsymbol{\eta}_0^{(i)}$ é utilizado como estimativa inicial para otimização não linear, convergindo para o ótimo local mais próximo $\boldsymbol{\eta}_i^*$. A solução final $\mathbf{u}_a^*$ corresponde ao vetor que produziu a maior norma dentre todas as $N_{\text{tentativas}}$ executadas.

3.3.1.6 Algoritmo Implementado (GRASP)

A implementação computacional do ataque de superamostragem, empregando a meta-heurística GRASP para a geração de pontos iniciais na etapa de otimização, encontra-se descrita no Apêndice A. Especificamente, o pseudocódigo completo do método C-GRASP aplicado ao sistema T-S MIMO e os procedimentos de construção da solução podem ser consultados na Seção A.1.

3.3.2 Formulação do Ataque via Operadores de *Koopman*

A teoria dos operadores de *Koopman* oferece uma abordagem alternativa para a análise de sistemas dinâmicos não lineares, representando a evolução do sistema em um espaço de funções observáveis em que a dinâmica é linear (KOOPMAN, 1931; MEZIĆ, 2005). Essa abordagem combinada com técnicas de identificação baseadas em dados como a decomposição em modos dinâmicos estendida (EDMD, do inglês *extended dynamic mode decomposition*) (WILLIAMS; KEVREKIDIS; ROWLEY, 2015), permite aproximar sistemas não lineares por modelos lineares de alta dimensão sem necessidade de conhecimento analítico da estrutura do sistema.

3.3.2.1 Sistema Original Não Linear

Considera-se o mesmo sistema não linear afim no controle descrito na Subseção 3.3.1.1. O operador de *Koopman* $\mathcal{K}$ é um operador linear de dimensão infinita que atua sobre funções observáveis. Uma função observável $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ é qualquer função escalar do estado do sistema, como posição, velocidade, energia ou combinações não lineares dessas variáveis. O operador é definido por sua ação de composição:

$$\mathcal{K}\psi(\mathbf{x}) = \psi(\mathbf{F}(\mathbf{x}, \mathbf{u})), \quad (3.62)$$

em que $\mathbf{F}$ representa o mapa de transição de estados do sistema. A propriedade fundamental é que, embora o sistema seja não linear, a evolução de observáveis é governada por dinâmica linear $\psi(\mathbf{x}(t+\Delta t)) = \mathcal{K}\psi(\mathbf{x}(t))$. Na prática, aproxima-se o operador infinito dimensional por uma matriz de dimensão finita $N \gg n$, definindo-se um dicionário de funções observáveis $\boldsymbol{\psi} : \mathbb{R}^n \rightarrow \mathbb{R}^N$:

$$\boldsymbol{\psi}(\mathbf{x}) = [\psi_1(\mathbf{x}) \ \psi_2(\mathbf{x}) \ \cdots \ \psi_N(\mathbf{x})]^\top, \quad (3.63)$$

representando o estado elevado $\mathbf{z} = \boldsymbol{\psi}(\mathbf{x}) \in \mathbb{R}^N$. Dicionários comuns incluem o polinomial $\boldsymbol{\psi}_{pol}(\mathbf{x}) = [1 \ x_1 \ x_2 \ \cdots \ x_1^2 \ x_1 x_2 \ \cdots \ x_n^d]^\top$, o trigonométrico $\boldsymbol{\psi}_{trig}(\mathbf{x}) = [1 \ \sin(x_1) \ \cos(x_1) \ \sin(x_2) \ \cos(x_2) \ \cdots]^\top$, e as funções de base radial (RBF, do inglês *radial basis function*):

$$\psi_i(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_i\|^2}{2\sigma^2}\right), \quad (3.64)$$

sendo $\mathbf{c}_i$ os centros e σ a largura da função.

3.3.2.2 Sistema Discreto com Taxa de Amostragem T_D

A decomposição em modos dinâmicos estendida é um método baseado em dados que identifica uma aproximação de dimensão finita do operador de *Koopman*. O método opera coletando um conjunto de M pares de *snapshots* $\{(\mathbf{x}_i, \mathbf{u}_i, \mathbf{x}_{i+1})\}_{i=1}^M$ do sistema com período de amostragem T_D , obtidos experimentalmente ou por simulação. Aplicando-se o

dicionário $\psi(\cdot)$ aos estados observados, constroem-se as matrizes de dados:

$$\mathbf{Z} = \begin{bmatrix} \psi(\mathbf{x}_1) & \psi(\mathbf{x}_2) & \cdots & \psi(\mathbf{x}_M) \end{bmatrix} \in \mathbb{R}^{N \times M}, \quad (3.65)$$

$$\mathbf{Z}' = \begin{bmatrix} \psi(\mathbf{x}_2) & \psi(\mathbf{x}_3) & \cdots & \psi(\mathbf{x}_{M+1}) \end{bmatrix} \in \mathbb{R}^{N \times M}, \quad (3.66)$$

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_M \end{bmatrix} \in \mathbb{R}^{m \times M}. \quad (3.67)$$

O método EDMD busca matrizes $\mathbf{A}_K \in \mathbb{R}^{N \times N}$ e $\mathbf{B}_K \in \mathbb{R}^{N \times m}$ que minimizem o resíduo entre o estado elevado previsto e o observado, utilizando a norma de *Frobenius*:

$$\min_{\mathbf{A}_K, \mathbf{B}_K} \|\mathbf{Z}' - (\mathbf{A}_K \mathbf{Z} + \mathbf{B}_K \mathbf{U})\|_F^2, \quad (3.68)$$

cujas soluções de mínimos quadrados é:

$$\begin{bmatrix} \mathbf{A}_K & \mathbf{B}_K \end{bmatrix} = \mathbf{Z}' \begin{bmatrix} \mathbf{Z} \\ \mathbf{U} \end{bmatrix}^\dagger, \quad (3.69)$$

sendo $(\cdot)^\dagger$ a pseudo-inversa de *Moore-Penrose*, computada via decomposição em valores singulares (SVD, do inglês *Singular Value Decomposition*). Quando a matriz de regressores $\mathbf{R} = \begin{bmatrix} \mathbf{Z}^\top & \mathbf{U}^\top \end{bmatrix}^\top \in \mathbb{R}^{(N+m) \times M}$ é mal-condicionada, emprega-se a regularização de *Tikhonov* (também conhecida como *Ridge Regression*), adicionando-se um termo de penalização à função objetivo (BRUNTON; KUTZ, 2019; KUTZ et al., 2016):

$$\min_{\mathbf{A}_K, \mathbf{B}_K} \|\mathbf{Z}' - (\mathbf{A}_K \mathbf{Z} + \mathbf{B}_K \mathbf{U})\|_F^2 + \lambda \left\| \begin{bmatrix} \mathbf{A}_K & \mathbf{B}_K \end{bmatrix} \right\|_F^2, \quad (3.70)$$

em que $\lambda > 0$ é o parâmetro de regularização. A solução regularizada é:

$$\begin{bmatrix} \mathbf{A}_K & \mathbf{B}_K \end{bmatrix} = \mathbf{Z}' \mathbf{R}^\top (\mathbf{R} \mathbf{R}^\top + \lambda \mathbf{I})^{-1}, \quad (3.71)$$

garantindo que a matriz seja sempre inversível. O parâmetro λ permite o compromisso entre fidelidade aos dados e estabilidade numérica: valores pequenos (tipicamente $\lambda \in [10^{-6}, 10^{-3}]$) estabilizam a solução sem distorção significativa, enquanto valores elevados produzem soluções suaves que podem não capturar adequadamente a dinâmica.

O modelo linear aproximado no espaço elevado é:

$$\mathbf{z}_{k+1} = \mathbf{A}_K \mathbf{z}_k + \mathbf{B}_K \mathbf{u}_k^*, \quad (3.72)$$

sendo $\mathbf{z}_k = \psi(\mathbf{x}_k) \in \mathbb{R}^N$ o estado elevado. Para o controlador, assume-se a lei de controle linear no espaço elevado, obtido por realimentação de estados:

$$\mathbf{u}_k^* = -\mathbf{K}_K \mathbf{z}_k, \quad (3.73)$$

em que $\mathbf{K}_K \in \mathbb{R}^{m \times N}$ é a matriz de ganho projetada no espaço de *Koopman* via LQR (do inglês *linear quadratic regulator*).

3.3.2.3 Identificação do Modelo Superamostrado

Existem duas abordagens para obter o modelo de *Koopman* superamostrado:

Abordagem 1 - Identificação Direta: Aplica-se o EDMD diretamente aos dados com período T_d , obtendo-se matrizes $\bar{\mathbf{A}}_{d,K} \in \mathbb{R}^{N \times N}$ e $\bar{\mathbf{B}}_{d,K} \in \mathbb{R}^{N \times m}$ que satisfazem:

$$\bar{\mathbf{z}}_{\ell+1} = \bar{\mathbf{A}}_{d,K} \bar{\mathbf{z}}_{\ell} + \bar{\mathbf{B}}_{d,K} \bar{\mathbf{u}}_{\ell}. \quad (3.74)$$

Abordagem 2 - Reamostragem: Se as matrizes $\mathbf{A}_K$ e $\mathbf{B}_K$ foram identificadas com período T_D e o sistema for aproximadamente linear no espaço elevado, pode-se usar exponenciação de matriz. Para sistemas em tempo contínuo, primeiro obtém-se a representação contínua:

$$\mathbf{A}_{K,cont} = \frac{1}{T_D} \log(\mathbf{A}_K), \quad \mathbf{B}_{K,cont} = (\mathbf{A}_K - \mathbf{I})^{-1} \mathbf{A}_{K,cont} \mathbf{B}_K, \quad (3.75)$$

e então discretiza-se com período T_d :

$$\bar{\mathbf{A}}_{d,K} = \exp(\mathbf{A}_{K,cont} T_d), \quad \bar{\mathbf{B}}_{d,K} = \mathbf{A}_{K,cont}^{-1} (\bar{\mathbf{A}}_{d,K} - \mathbf{I}) \mathbf{B}_{K,cont}. \quad (3.76)$$

3.3.2.4 Condição de Indetectabilidade

Considerando o vetor de ataque, descrito na Equação (3.51). A evolução do estado elevado ao longo de r passos superamostrados é:

$$\bar{\mathbf{z}}_{\ell+r} = [\bar{\mathbf{A}}_{d,K}]^r \bar{\mathbf{z}}_{\ell} + \sum_{j=0}^{r-1} [\bar{\mathbf{A}}_{d,K}]^{r-1-j} \bar{\mathbf{B}}_{d,K} \bar{\mathbf{u}}_{\ell+j}. \quad (3.77)$$

Definindo as matrizes agregadas:

$$\bar{\Theta}_K = [\bar{\mathbf{A}}_{d,K}]^r \in \mathbb{R}^{N \times N}, \quad (3.78)$$

$$\bar{\Gamma}_K = \begin{bmatrix} [\bar{\mathbf{A}}_{d,K}]^{r-1} \bar{\mathbf{B}}_{d,K} & [\bar{\mathbf{A}}_{d,K}]^{r-2} \bar{\mathbf{B}}_{d,K} & \cdots & \bar{\mathbf{A}}_{d,K} \bar{\mathbf{B}}_{d,K} & \bar{\mathbf{B}}_{d,K} \end{bmatrix} \in \mathbb{R}^{N \times rm}. \quad (3.79)$$

A condição de indetectabilidade no espaço de *Koopman* completo (N dimensões) é:

$$\bar{\Gamma}_K \mathbf{u}_a = \mathbf{d}_k^{K,full}, \quad (3.80)$$

em que o vetor de desvio no espaço elevado é:

$$\mathbf{d}_k^{K,full} = \psi(\mathbf{x}_{k+1}^{nom}) - \bar{\Theta}_K \psi(\mathbf{x}_k) = \mathbf{z}_{k+1}^{nom} - \bar{\Theta}_K \mathbf{z}_k \in \mathbb{R}^N, \quad (3.81)$$

com $\mathbf{z}_{k+1}^{nom} = \mathbf{A}_K \mathbf{z}_k + \mathbf{B}_K \mathbf{u}_k^*$.

3.3.2.5 Projeção para o Espaço Físico

O sistema descrito pela Equação (3.80) possui N equações, o que pode ser restritivo dado que $N \gg n$. Na prática, apenas as componentes correspondentes aos n estados físicos originais precisam satisfazer a condição de furtividade.

Seja $\mathbf{C}_{phys} \in \mathbb{R}^{n \times N}$ a matriz de projeção que seleciona as n componentes do dicionário correspondentes aos estados físicos. Tipicamente, quando o dicionário é construído de modo que $\psi_i(\mathbf{x}) = x_i$ para $i = 1, \dots, n$, tem-se:

$$\mathbf{C}_{phys} = \begin{bmatrix} \mathbf{I}_{n \times n} & \mathbf{0}_{n \times (N-n)} \end{bmatrix}. \quad (3.82)$$

A condição de furtividade projetada é:

$$\mathbf{C}_{phys} \bar{\mathbf{\Gamma}}_K \mathbf{u}_a = \mathbf{C}_{phys} \mathbf{d}_k^{K,full}. \quad (3.83)$$

Definindo as matrizes efetivas:

$$\mathbf{c}_r^K = \mathbf{C}_{phys} \bar{\mathbf{\Gamma}}_K \in \mathbb{R}^{n \times rm}, \quad (3.84)$$

$$\mathbf{d}_k^K = \mathbf{C}_{phys} \mathbf{d}_k^{K,full} \in \mathbb{R}^n. \quad (3.85)$$

A condição de indetectabilidade torna-se:

$$\mathbf{c}_r^K \mathbf{u}_a = \mathbf{d}_k^K. \quad (3.86)$$

Essa formulação recupera a estrutura do problema linear com n restrições de igualdade e rm variáveis de decisão, análoga à formulação T-S.

3.3.2.6 Problema de Otimização do Ataque

O problema de otimização do ataque de superamostragem via Operadores de *Koopman* mantém a estrutura do problema expressa pela Equação (3.13), substituindo-se:

- $\mathbf{c}_r$ por $\mathbf{c}_r^K$;
- $\mathbf{d}_k$ por $\mathbf{d}_k^K$.

3.3.2.7 Métodos de Solução

Os quatro métodos de solução apresentados para a formulação T-S (Subseção 3.3.1.5) são diretamente aplicáveis à formulação via Operadores de *Koopman*, uma vez que ambas resultam em problemas de programação quadrática com restrições lineares de igualdade e desigualdade.

3.3.2.8 Algoritmo Implementado (GRASP)

O pseudocódigo e os procedimentos específicos desta abordagem (EDMD-GRASP) estão descritos na Seção A.2. Ressalta-se que a estrutura geral preserva analogia com a formulação T-S, distinguindo-se pela etapa de identificação *offline* baseada em dados e pela ausência do cálculo *online* de funções de pertinência.

3.3.2.9 Comparação entre as Formulações T-S e *Koopman*

Embora ambas as abordagens resultem no mesmo problema de otimização quadrática, elas diferem nos requisitos de modelagem, na complexidade computacional e nas hipóteses necessárias para sua aplicação. A Tabela 1 apresenta um sumário comparativo entre as formulações T-S e *Koopman*.

Tabela 1 – Comparação entre formulações *Takagi-Sugeno* e Operadores de *Koopman* para ataque de superamostragem.

Critério	<i>Takagi-Sugeno</i> (T-S)	Operadores de <i>Koopman</i>
1. Fundamentos Teóricos		
Modelo	Interpolação convexa de modelos lineares locais	Elevação para espaço de observáveis com dinâmica linear
Tipo de modelo	Caixa-cinza (estrutura física preservada)	Caixa-preta (baseado em dados)
Espaço de estados	Original: $\mathbb{R}^n$	Elevado: $\mathbb{R}^N$ ($N \gg n$)
Representação	$\sum_{i=1}^{2^p} h_i(\mathbf{z}_k) [\mathbf{A}_i \mathbf{x}_\ell + \mathbf{B}_i \mathbf{u}_\ell]$	$\mathbf{z}_{k+1} = \mathbf{A}_K \mathbf{z}_k + \mathbf{B}_K \mathbf{u}_k,$ $\mathbf{z} = \boldsymbol{\psi}(\mathbf{x})$
Dependência temporal	Variante (h_i muda a cada k)	Invariante (matrizes fixas)
Interpretabilidade	Alta (subsistemas físicos)	Baixa (espaço abstrato)
2. Construção do Modelo (Fase <i>Offline</i>)		
Entrada requerida	Modelo analítico: $\mathbf{f}(\mathbf{x})$, $\mathbf{G}(\mathbf{x})$	Dados: $\{(\mathbf{x}_i, \mathbf{u}_i, \mathbf{x}_{i+1})\}_{i=1}^M$
Processamento	• Identificar p não linearidades • Calcular limites setoriais • Gerar 2^p subsistemas • Discretizar (ZOH)	• Escolher dicionário $\boldsymbol{\psi}$ • Aplicar aos dados • EDMD: resolver mín. quadrados • Pré-computar $\bar{\boldsymbol{\Theta}}_K, \bar{\boldsymbol{\Gamma}}_K$

Continua na próxima página...

Continuação da página anterior...

Critério	<i>Takagi-Sugeno</i> (T-S)	Operadores de <i>Koopman</i>
Complexidade	$\mathcal{O}(2^p \cdot n^3)$	$\mathcal{O}(MN^2 + N^3)$
Escalabilidade	Explosão exponencial em p	Crescimento cúbico em N
3. Execução <i>Online</i> (a cada kT_D)		
Operações principais	1. Calcular $h_i(\mathbf{z}_k)$ (2^p funções) 2. Agregar: $\bar{\mathbf{A}}_d(h_k) = \sum h_i \bar{\mathbf{A}}_{d,i}$ 3. Potenciar: $[\bar{\mathbf{A}}_d]^j, j = 1, \dots, r$ 4. Construir $\mathcal{C}_r(h_k)$	1. Avaliar $\mathbf{z}_k = \psi(\mathbf{x}_k)$ 2. Usar matrizes pré-computadas 3. Construir $\mathcal{C}_r^K$ (já calculada)
Custo computacional	$\mathcal{O}(2^p n^2 + r n^3)$	$\mathcal{O}(N \cdot c_\psi + N^2)$
Dependência de r	Linear (potenciações)	Nenhuma (pré-computado)
Vantagem quando	$p \leq 4, n \leq 10, r$ pequeno ¹	$p \geq 5, r \geq 10, N$ moderado
4. Hipóteses e Acurácia		
Hipótese	$h_i(\bar{\mathbf{z}}_{\ell+m}) \approx h_i(\mathbf{z}_k), \forall m \leq r$	$\psi(\mathbf{x})$ rico o suficiente para capturar dinâmica
Acurácia depende de	Cobertura setorial, validade da Hipótese 1	Escolha de ψ , qualidade/quantidade de dados
5. Aplicabilidade Prática		
Cenário ideal	• Modelo analítico disponível • Poucas não linearidades ($p \leq 4$) • Interpretabilidade	• Sistema caixa-preta • Muitas não linearidades ($p > 5$) • Dados em grandes quantidades

Continua na próxima página...

¹ Os limiares indicados são de natureza heurística e foram estabelecidos com base em testes experimentais realizados pelo autor ao longo do desenvolvimento deste trabalho.

Continuação da página anterior...

Critério	<i>Takagi-Sugeno</i> (T-S)	Operadores de <i>Koopman</i>
Limitações	• Inviável para $p > 6$ (64+ regras) • Custo cresce com r • Requer modelo analítico	• Incapacidade de previsão correta • Sensível a ruído • Interpretabilidade
6. Problema de Otimização		
Estrutura	Idêntica: $\max \ \mathbf{u}_a\ ^2$ s.a. $\mathcal{C}_r \mathbf{u}_a = \mathbf{d}_k$, $\underline{\mathbf{u}} \leq \bar{\mathbf{u}}_{\ell+j} \leq \bar{\mathbf{u}}$	
Diferença	$\mathcal{C}_r(h_k)$ varia com h_k	$\mathcal{C}_r^K$ fixo (pré-computado)
Métodos de solução	Vértices, SDP, Espaço Nulo, <i>Multi-Start</i> (GRASP)	

4 Resultados e Procedimentos Experimentais

Este capítulo apresenta os resultados experimentais e computacionais obtidos ao longo do trabalho, organizados em três cenários progressivos. O Cenário 1 valida o ataque de superamostragem em sistemas lineares por meio de simulações numéricas, incluindo problemas de regulação, rastreamento e extensão para sistemas MIMO. O Cenário 2 estende a análise para sistemas não lineares, utilizando modelagem *Takagi-Sugeno* e operadores de *Koopman* aplicados a um motor DC com saturação magnética. Por fim, o Cenário 3 adota a perspectiva do atacante para demonstrar a viabilidade do ataque em uma planta física real, incluindo as etapas de reconhecimento, preparação e execução.

4.1 Cenário 1: Validando o Ataque de Superamostragem Computacionalmente em Sistemas Lineares

Para avaliar o impacto do ataque de superamostragem e validar sua eficácia, foram realizados experimentos numéricos considerando dois tipos de problemas de controle: regulação e rastreamento. Em ambos os casos, o sistema foi submetido a condições com presença de ruído de processo para simular um cenário mais realista de operação.

4.1.1 Descrição do Sistema e Configuração Experimental

O sistema considerado é um integrador duplo linear de segunda ordem com entrada escalar (SISO), descrito pelas seguintes matrizes na forma contínua:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 1 & 0 \end{bmatrix}. \quad (4.1)$$

Esse sistema representa uma dinâmica física em que $\dot{x}_1 = x_2$ (a derivada da posição é a velocidade) e $\dot{x}_2 = u$ (a derivada da velocidade é a ação de controle). Para a presença de ruído de processo, foi considerado o vetor:

$$\mathbf{b}_w = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad (4.2)$$

com ruído gaussiano de média nula e desvio padrão $\sigma = 0,2$. O ruído foi implementado como $w = \mathbf{b}_w \cdot \sigma \cdot \mathcal{N}(0,1)$, em que $\mathcal{N}(0,1)$ representa uma distribuição normal padrão.

O sistema contínuo foi discretizado utilizando o método ZOH (FRANKLIN et al., 1998) (função `c2d` do MATLAB®). Com o período de amostragem original do sistema, fixado em $T_D = 0,1$ s.

Para todos os experimentos, foram adotadas as seguintes condições iniciais:

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad (4.3)$$

com limites de saturação para a ação de controle no intervalo $[\underline{u}, \bar{u}] = [-15, 15]$. O tempo total de simulação foi de 15 s. Para a implementação do ataque, a taxa de amostragem foi aumentada por um fator $r = 4$, resultando em um novo período de amostragem $T_d = T_D/r = 0,025$ s.

4.1.2 Problema de Regulação

O problema de regulação consiste em estabilizar o sistema no ponto de equilíbrio $\mathbf{x}_{\text{ref}} = \mathbf{0}$, partindo das condições iniciais especificadas. Para esse problema, foi empregado o controlador linear quadrático discreto (DLQR, do inglês *discrete linear quadratic regulator*) (OGATA, 1995) (função `dlqr` do MATLAB®).

As matrizes de ponderação utilizadas no projeto do controlador DLQR foram obtidas através de testes experimentais, sendo selecionadas por apresentarem os melhores resultados em termos de rapidez de resposta e desempenho global do sistema:

$$\mathbf{Q} = \mathbf{I}_2, \quad R = 1, \quad (4.4)$$

em que $\mathbf{Q}$ penaliza o desvio dos estados em relação ao ponto de equilíbrio e R penaliza o uso da ação de controle.

Para o problema de regulação, o ganho obtido foi:

$$\mathbf{k}_{\text{reg}} = [0,9171 \quad 1,6356]^\top. \quad (4.5)$$

As Figuras 7 e 8 apresentam a evolução temporal das variáveis de estado x_1 e x_2 para os casos com e sem ataque de superamostragem na presença de ruído de processo. Em azul sólido é mostrado o comportamento do sistema sem ataque operando com $T_D = 0,1$ s, enquanto em vermelho sólido é apresentado o sistema sob ataque com $T_d = 0,025$ s.

As Figuras 9 e 10 mostram que o comportamento das variáveis de estado sob ataque é similar ao nominal nos instantes de amostragem originais. Embora a presença de ruído impeça uma coincidência perfeita entre os estados, essa diferença é irrelevante para a furtividade do ataque. Como o ruído é branco e de média nula, qualquer desvio é absorvido pela dinâmica estocástica, resultando em um resíduo do filtro de *Kalman* que permanece branco, tornando o ataque indistinguível.

A Figura 11 apresenta a ação de controle para ambos os cenários. Nota-se que, sob ataque, a ação de controle exibe oscilações significativas e atinge frequentemente os limites de saturação impostos (± 15). Esse comportamento é resultado direto do problema de

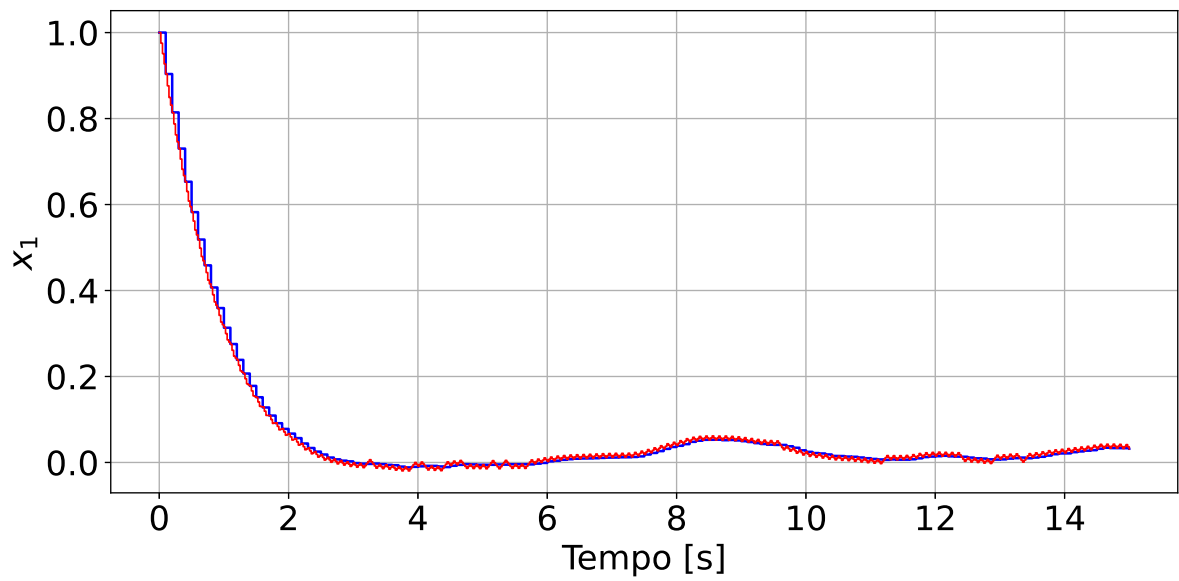


Figura 7 – Evolução temporal da variável de estado x_1 para o problema de regulação com presença de ruído. Em azul sólido, o sistema sem ataque com $T_D = 0,1$ s; em vermelho sólido, o sistema sob ataque de superamostragem com $T_d = 0,025$ s.

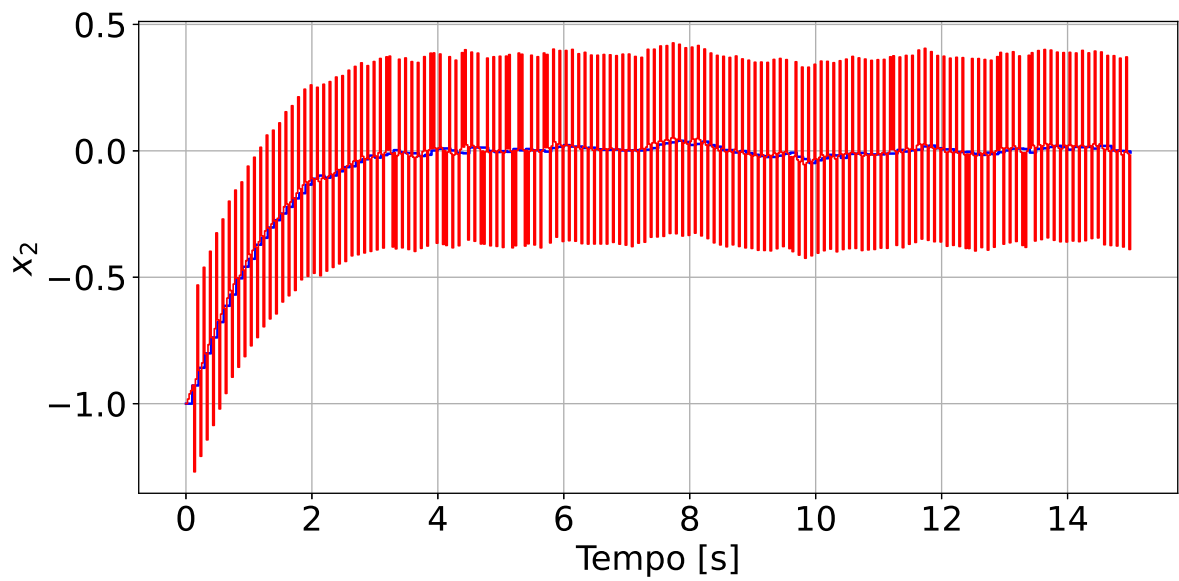


Figura 8 – Evolução temporal da variável de estado x_2 para o problema de regulação com presença de ruído. Para mais informações, ver legenda da Figura 7.

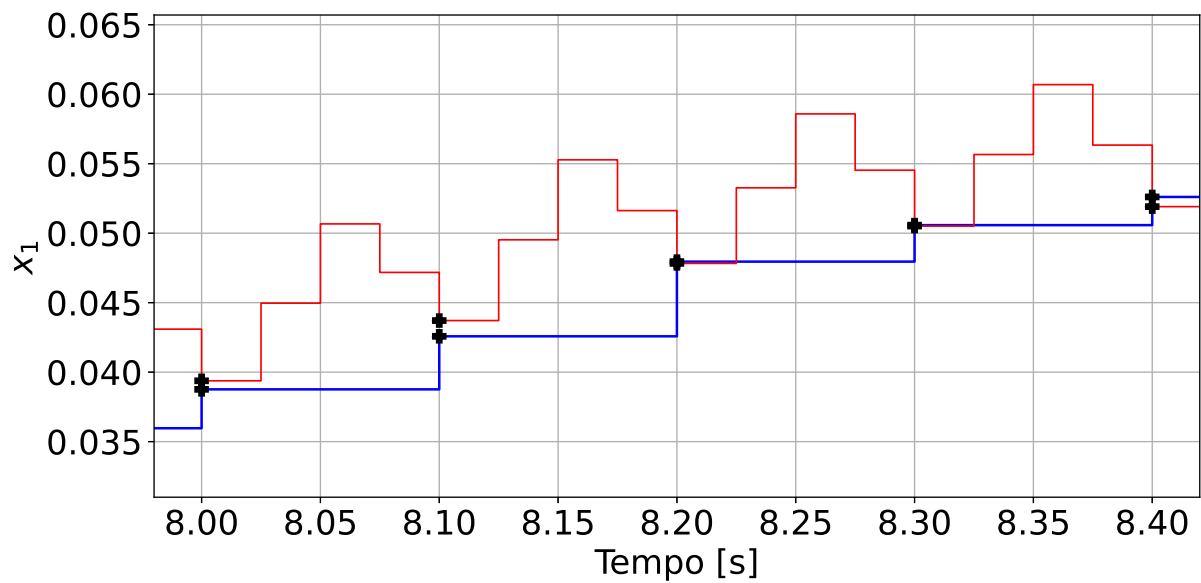


Figura 9 – *Zoom* da evolução temporal da variável de estado x_1 para o problema de regulação com presença de ruído. As cruzes indicam os instantes de amostragem original. Para mais informações, ver legenda da Figura 7.

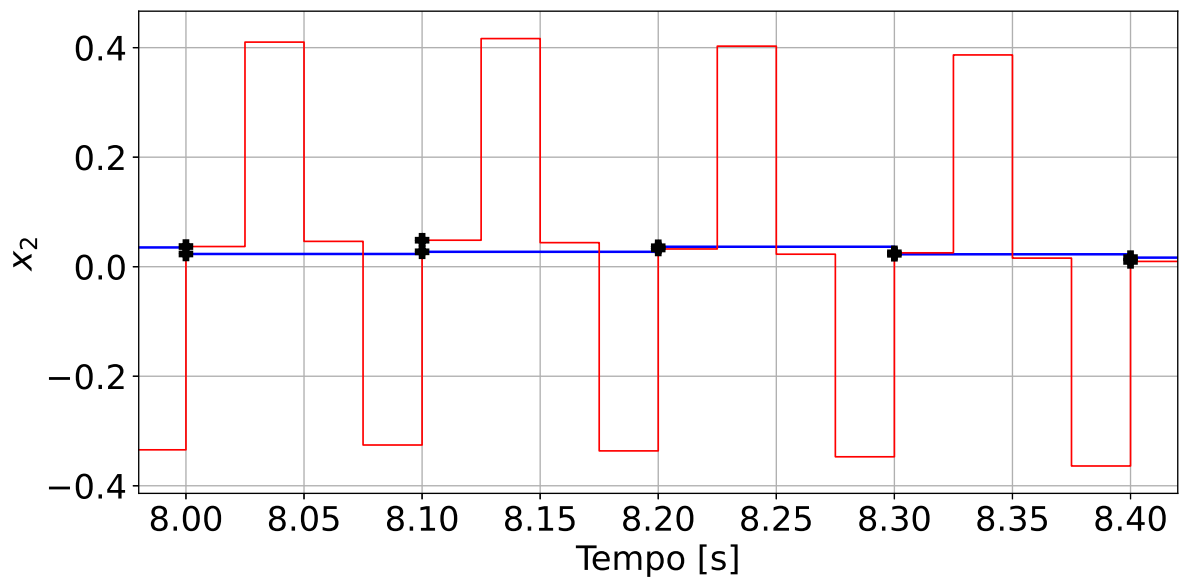


Figura 10 – *Zoom* da evolução temporal da variável de estado x_2 para o problema de regulação com presença de ruído. As cruzes indicam os instantes de amostragem original. Para mais informações, ver legenda da Figura 7.

otimização formulado para o ataque, que busca maximizar o consumo energético enquanto mantém a saída inalterada nos instantes de amostragem originais.

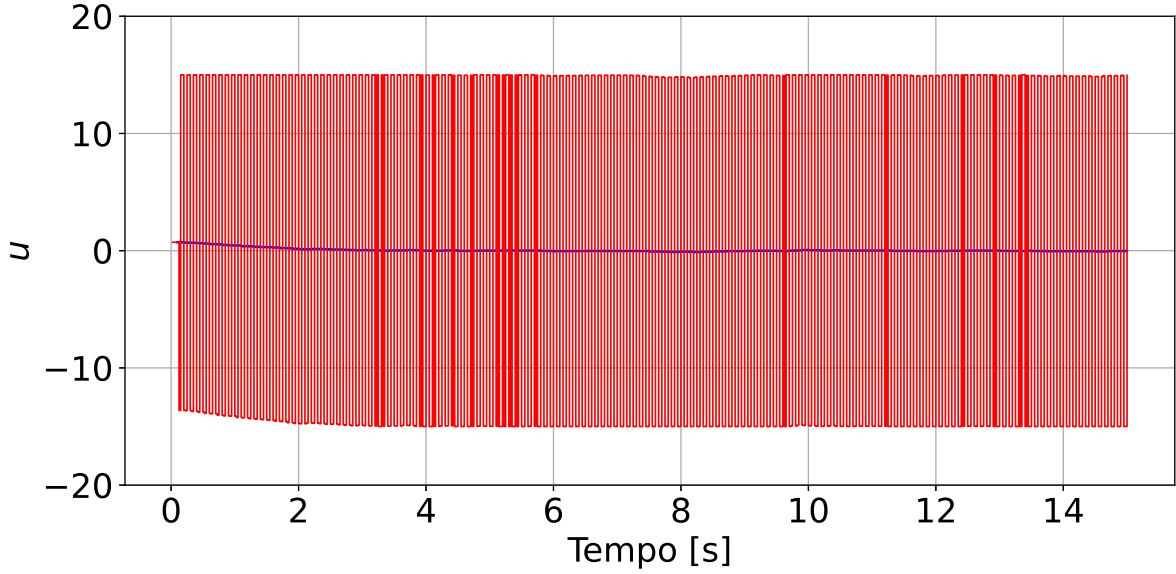


Figura 11 – Ação de controle para o problema de regulação com presença de ruído. Em azul sólido, o sistema sem ataque com $T_D = 0,1$ s; em vermelho sólido, o sistema sob ataque de superamostragem com $T_d = 0,025$ s.

A análise do consumo energético, quantificado pela métrica E_{RMS} definida na Equação (3.14), mostra o impacto do ataque. Para o problema de regulação, a energia consumida pelo sistema sem ataque foi de $E_{RMS_{sem}} = 0,1739$, enquanto no sistema sob ataque esse valor aumentou para $E_{RMS_{com}} = 14,8695$. Isso representa, de acordo com a Equação (3.15), um incremento percentual de:

$$\Delta E(\%) = \frac{E_{RMS}^{ataque} - E_{RMS}^{normal}}{E_{RMS}^{normal}} \times 100 = \frac{14,8695 - 0,1739}{0,1739} \times 100\% = 8451\%. \quad (4.6)$$

Esse incremento demonstra que o ataque pode causar um aumento de mais de 84 vezes no consumo energético do sistema, levando potencialmente ao desgaste dos atuadores e comprometendo a viabilidade operacional do sistema em longo prazo.

Além do consumo energético, outro aspecto a ser avaliado é o desgaste imposto aos atuadores devido às variações na ação de controle. Para quantificar esse efeito, utiliza-se a Equação (3.16).

Para o problema de regulação, o índice TV do sistema sem ataque foi de 2,25, enquanto sob ataque de superamostragem esse valor aumentou para 9418,19, representando um incremento de 418676%. Esse aumento no índice de variação total evidencia que o ataque não apenas eleva o consumo energético, mas também impõe ciclos de trabalho muito mais severos aos atuadores. As oscilações de alta frequência, aceleram o desgaste mecânico dos componentes, podendo levar à fadiga prematura, aumento nos custos de

manutenção e, em casos extremos, falha completa do atuador, comprometendo a segurança e confiabilidade do sistema.

4.1.3 Problema de Rastreamento

O problema de rastreamento consiste em fazer com que os estados do sistema converjam para uma referência desejada não nula. Para este experimento, foi considerado o sinal de referência $\mathbf{x}_{\text{ref}}(k) = [2, 0]^\top$ para $k \geq 0$.

As matrizes de ponderação, foram definidas como:

$$\mathbf{Q} = \mathbf{I}_2, \quad R = 1. \quad (4.7)$$

O projeto do controlador para rastreamento requer a construção de um sistema aumentado que incorpora a dinâmica do erro integral. Considerando o sistema discretizado original descrito por $\mathbf{x}_{k+1} = \mathbf{A}_D \mathbf{x}_k + \mathbf{b}_D \mathbf{u}_k$ com saída $\mathbf{y}_k = \mathbf{c} \mathbf{x}_k$, define-se o estado aumentado como:

$$\mathbf{x}_{\text{aug},k} = \begin{bmatrix} \mathbf{x}_k \\ \boldsymbol{\xi}_k \end{bmatrix}, \quad (4.8)$$

em que $\boldsymbol{\xi}_k$ representa o estado integrador definido por $\boldsymbol{\xi}_{k+1} = \boldsymbol{\xi}_k + (\mathbf{r} - \mathbf{y}_k)$, com $\mathbf{r}$ sendo a referência desejada. O sistema aumentado é então descrito por:

$$\mathbf{x}_{\text{aug},k+1} = \mathbf{A}_{\text{aug}} \mathbf{x}_{\text{aug},k} + \mathbf{b}_{\text{aug}} \mathbf{u}_k, \quad (4.9)$$

em que as matrizes aumentadas são construídas como:

$$\mathbf{A}_{\text{aug}} = \begin{bmatrix} \mathbf{A}_D & \mathbf{0} \\ -\mathbf{C} & \mathbf{I} \end{bmatrix}, \quad \mathbf{b}_{\text{aug}} = \begin{bmatrix} \mathbf{b}_D \\ \mathbf{0} \end{bmatrix}, \quad (4.10)$$

sendo $\mathbf{I}$ a matriz identidade de dimensão adequada. A matriz de ponderação aumentada para o problema de otimização DLQR é definida como:

$$\mathbf{Q}_{\text{aug}} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{\text{int}} \end{bmatrix}, \quad \text{com } \mathbf{Q}_{\text{int}} = 10 \cdot \mathbf{I}_2, \quad (4.11)$$

em que $\mathbf{Q}_{\text{int}}$ penaliza o erro integral dos estados. Essa matriz apresenta valores mais elevados para garantir que o erro de rastreamento seja eliminado em regime permanente.

O ganho de controle aumentado $\mathbf{k}_{\text{aug}}$ é então obtido resolvendo a equação algébrica de *Riccati* discreta para o sistema aumentado, com matrizes $\mathbf{A}_{\text{aug}}$, $\mathbf{b}_{\text{aug}}$, $\mathbf{Q}_{\text{aug}}$ e R . A lei de controle resultante é dada por:

$$\mathbf{u}_k = -\mathbf{k}_{\text{aug}} \mathbf{x}_{\text{aug},k} = -\mathbf{k}_x \mathbf{x}_k - \mathbf{k}_i \boldsymbol{\xi}_k, \quad (4.12)$$

em que $\mathbf{k}_{\text{aug}} = [\mathbf{k}_x \quad \mathbf{k}_i]$ é particionado de acordo com as componentes do estado aumentado. Essa formulação garante erro nulo em regime permanente para referências constantes, uma vez que a ação integral acumula o erro ao longo do tempo até eliminá-lo completamente.

Resolvendo a equação do DLQR com as devidas matrizes, foram encontrados os ganhos:

$$\mathbf{k}_{\text{aug}} = \begin{bmatrix} 27,3974 & 7,2241 & -2,0113 & -1,6727 \times 10^{-7} \end{bmatrix}^T. \quad (4.13)$$

As Figuras 12 e 13 apresentam a evolução temporal das variáveis de estado para o problema de rastreamento.

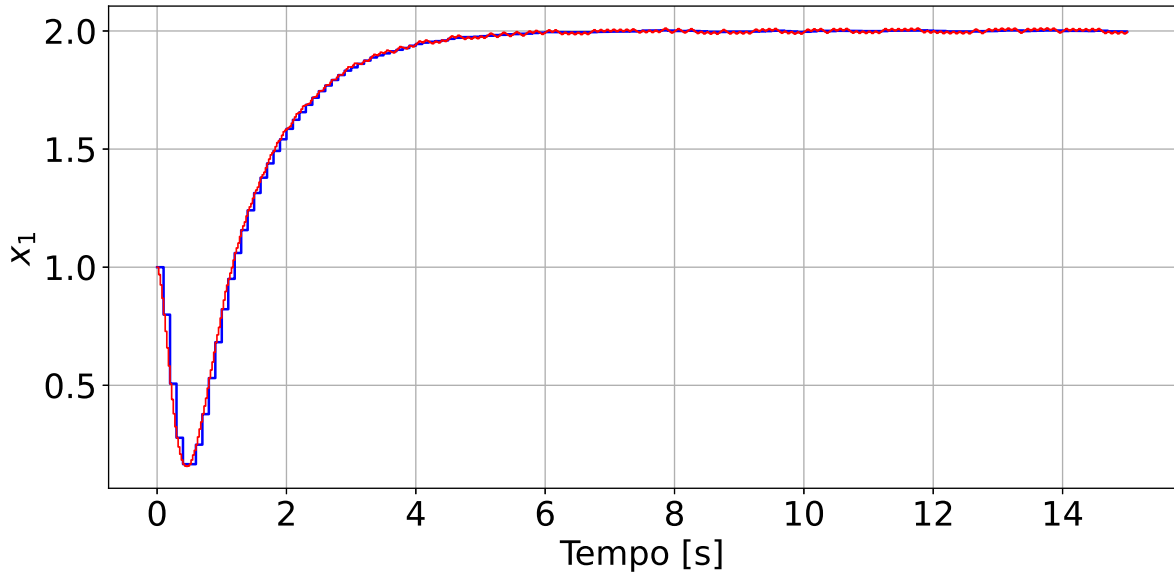


Figura 12 – Evolução temporal da variável de estado x_1 para o problema de rastreamento com presença de ruído. Para mais informações, ver legenda da Figura 7.

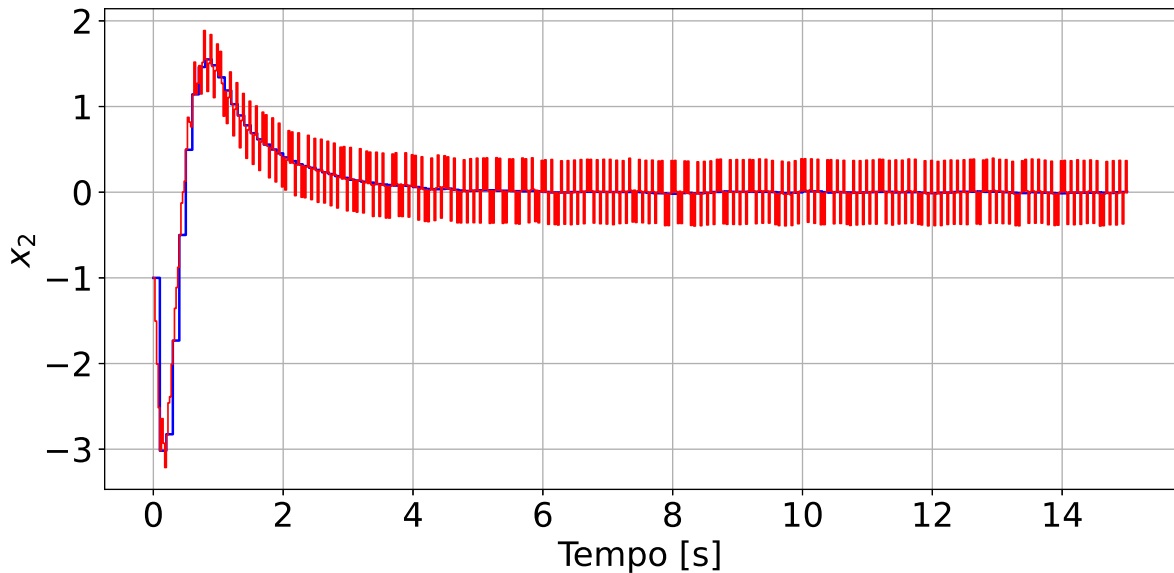


Figura 13 – Evolução temporal da variável de estado x_2 para o problema de rastreamento com presença de ruído. Para mais informações, ver legenda da Figura 8.

Conforme observado nas Figuras 14 e 15, a indetectabilidade do ataque continua no problema de rastreamento. Embora o ruído impeça a coincidência exata dos estados

em nT_D .

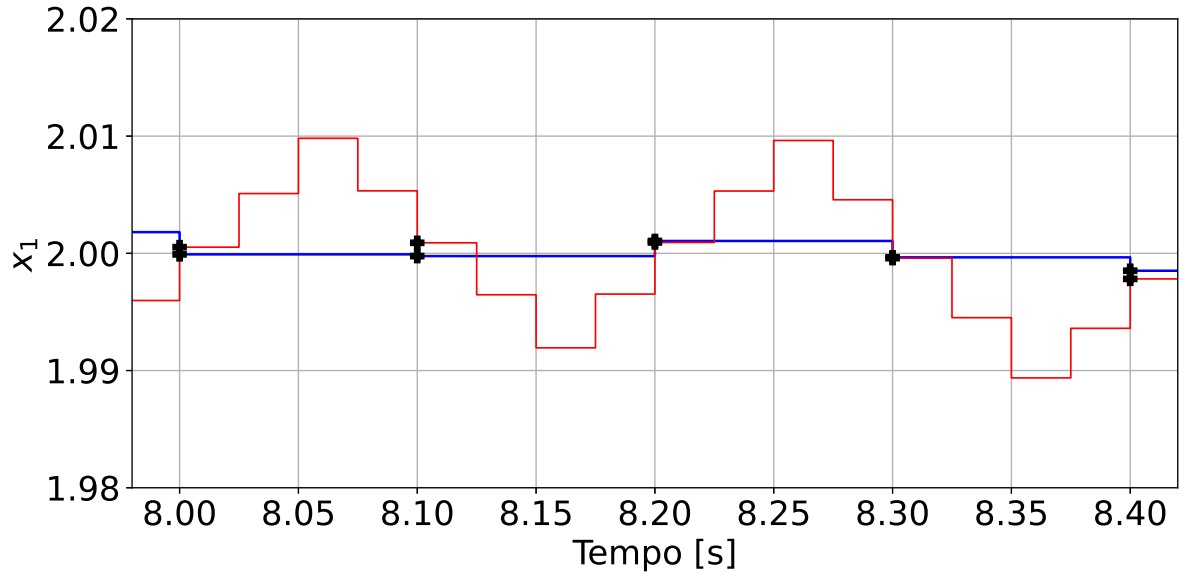


Figura 14 – *Zoom* da evolução temporal da variável de estado x_1 para o problema de rastreamento com presença de ruído. Para mais informações, ver legenda da Figura 9.

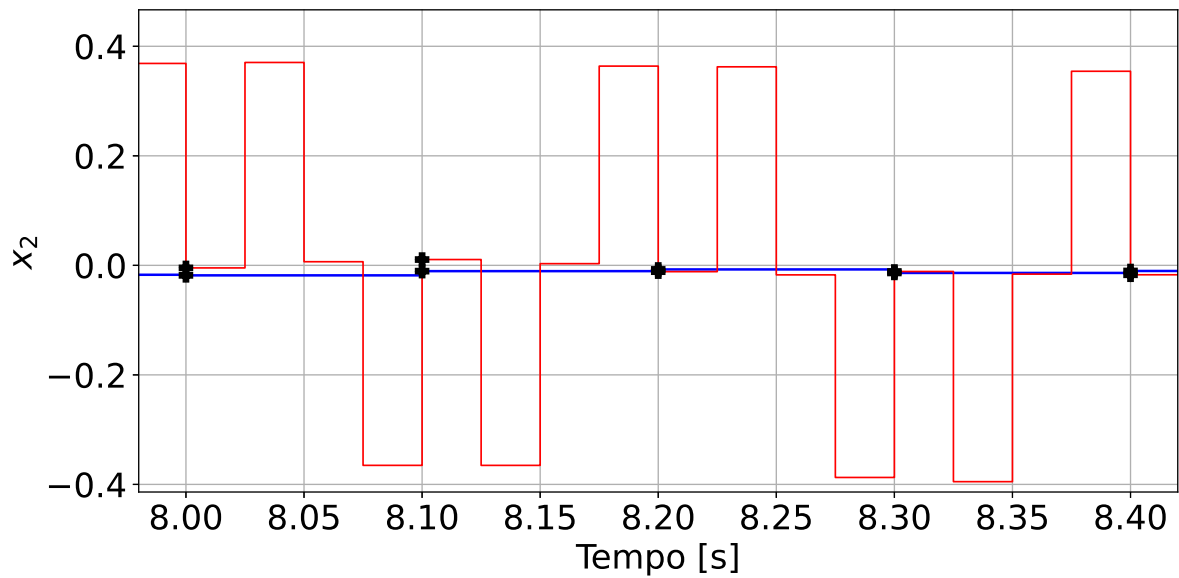


Figura 15 – *Zoom* da evolução temporal da variável de estado x_2 para o problema de rastreamento com presença de ruído. Para mais informações, ver legenda da Figura 10.

A Figura 16 ilustra a ação de controle para o problema de rastreamento. Similarmente ao caso de regulação, o sistema sob ataque apresenta oscilações acentuadas e saturação frequente da ação de controle.

Para o problema de rastreamento, a energia consumida pelo sistema sem ataque foi de $E_{RMSsem} = 2,3819$, enquanto sob ataque aumentou para $E_{RMScom} = 14,7753$. O

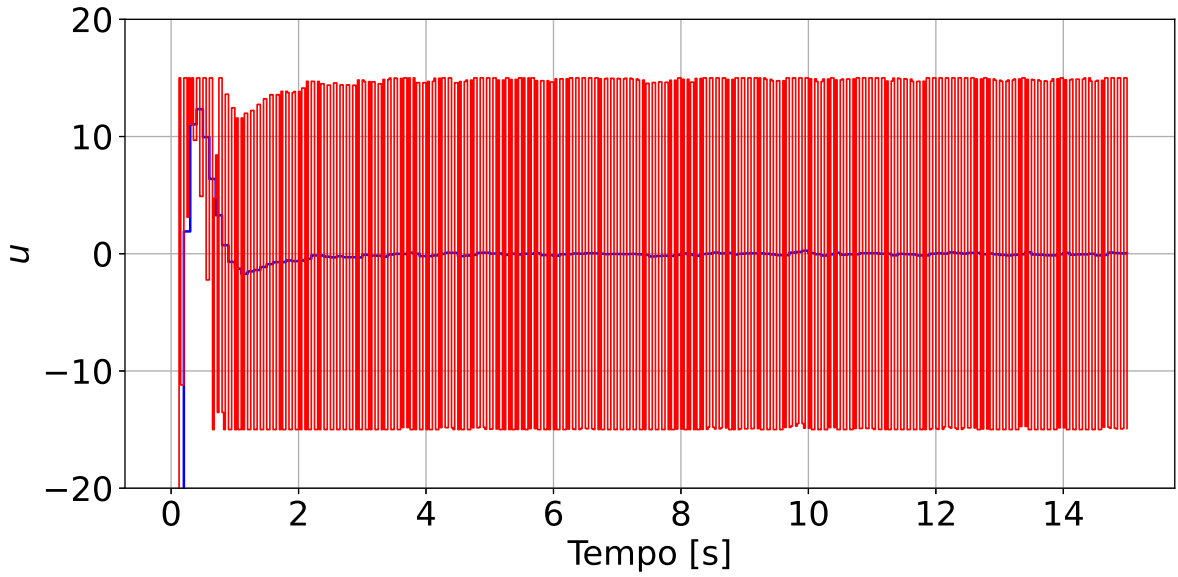


Figura 16 – Ação de controle para o problema de rastreamento com presença de ruído. Para mais informações, ver legenda da Figura 11.

incremento percentual nesse caso é:

$$\Delta E(\%) = \frac{14,7753 - 2,3819}{2,3819} \times 100\% = 520,3\%. \quad (4.14)$$

Similarmente ao caso de regulação, o desgaste dos atuadores constitui uma preocupação no problema de rastreamento. Utilizando o índice de variação total, observou-se que o sistema sem ataque apresentou $TV = 57,25$, ao passo que sob ataque de superamostragem esse valor elevou-se para 10292,39, correspondendo a um incremento de 17876,67%.

Embora o incremento percentual do índice TV seja menor ao observado no problema de regulação, o valor absoluto ainda representa um aumento na atividade do atuador.

4.1.4 Influência do Fator de Superamostragem

Para avaliar o impacto de diferentes fatores de superamostragem tanto no consumo energético quanto no desgaste dos atuadores, foram realizados experimentos variando o fator r de aumento da taxa de amostragem na faixa de 4 a 128. As Tabelas 2 e 3 apresentam os resultados obtidos para os problemas de regulação e rastreamento, respectivamente, considerando a presença de ruído de processo, reportando em cada caso as métricas E_{sem} , E_{com} , TV_{sem} e TV_{com} , bem como os respectivos incrementos percentuais.

Os resultados apresentados nas Tabelas 2 e 3 revelam características importantes do ataque de superamostragem sob duas perspectivas complementares: o consumo energético e o desgaste do atuador.

Em relação ao consumo energético, observa-se que E_{sem} permanece praticamente constante independentemente do fator de superamostragem considerado, apresentando

Tabela 2 – Energia consumida e índice de desgaste do atuador para diferentes fatores de superamostragem no problema de regulação com ruído ($\sigma = 0,2$).

Fator r	E_{sem}	E_{com}	Incremento E (%)	TV_{sem}	TV_{com}	Incremento TV (%)
4	0,1739	14,8695	8451,6	2,2490	9418	418676
8	0,1707	14,7244	8527,5	1,8268	26528	1452064
16	0,1686	14,1396	8287,5	1,4946	58569	3918727
32	0,1679	14,6296	8612,3	1,2960	88155	6802120
64	0,1690	14,8100	8664,7	1,1087	182613	16470566
128	0,1703	14,8820	8638,8	0,9773	331546	33925821

Tabela 3 – Energia consumida e índice de desgaste do atuador para diferentes fatores de superamostragem no problema de rastreamento com ruído ($\sigma = 0,2$).

Fator r	E_{sem}	E_{com}	Incremento E (%)	TV_{sem}	TV_{com}	Incremento TV (%)
4	2,3819	14,7753	520,3	57,2542	10292	17876
8	2,3758	14,7843	522,2	54,8185	26886	48945
16	2,3741	14,5386	512,4	52,8295	59356	112254
32	2,3754	14,7109	519,4	51,8914	114137	219854
64	2,3737	14,9036	527,9	50,4246	203735	403939
128	2,3734	14,9816	531,2	49,5062	348877	704614

variações mínimas devido apenas às flutuações estocásticas do ruído de processo. Para o problema de regulação, E_{sem} varia entre 0,1679 e 0,1739, enquanto para o rastreamento oscila entre 2,3734 e 2,3819, representando variações inferiores a 3,5% e 0,4%, respectivamente. Isso confirma que o fator de superamostragem utilizado pelo atacante não afeta o desempenho do sistema nominal. Por outro lado, E_{com} apresenta estabilidade em torno de valores próximos a 14,5–14,9 para ambos os problemas, independentemente do fator de superamostragem empregado. Essa consistência é uma consequência direta da formulação do problema de otimização do ataque, que busca maximizar o consumo energético sujeito às restrições físicas de saturação ($[-15, 15]$) e à condição de indetectabilidade. Uma vez que esses limites e a dinâmica do sistema permanecem inalterados, o algoritmo de otimização converge para soluções que exploram de forma similar o espaço admissível de controle, resultando em consumos energéticos comparáveis.

A diferença significativa nos incrementos percentuais de energia entre os dois problemas decorre da diferença no consumo energético nominal: o problema de regulação, partindo de $\mathbf{x}_0 = [1, -1]^\top$ e convergindo para a origem, requer esforço de controle significativamente menor ($E_{\text{sem}} \approx 0,17$) comparado ao rastreamento ($E_{\text{sem}} \approx 2,38$). Consequentemente, embora o impacto absoluto do ataque seja similar em ambos os casos, o impacto relativo é expressivamente maior para a regulação, com incrementos entre 8287,5% e 8664,7%, contra 512,4% a 531,2% no rastreamento.

Em relação ao índice de desgaste do atuador, observa-se um comportamento qualitativamente distinto do consumo energético: o TV_{com} cresce monotonicamente com o fator de superamostragem r em ambos os problemas. Para a regulação, TV_{com} varia de

9418 ($r = 4$) a 331546 ($r = 128$), representando um aumento de mais de 35 vezes entre os extremos testados, com incrementos percentuais que crescem de 418676% a 33925821%. Para o rastreamento, o crescimento é análogo, porém com valores absolutos maiores devido ao maior esforço de controle nominal: TV_{com} varia de 10292 ($r = 4$) a 348877 ($r = 128$), com incrementos de 17876% a 704614%. Esse comportamento é fisicamente esperado: fatores de superamostragem maiores implicam ações de controle aplicadas em intervalos menores, aumentando o número de transições bruscas por unidade de tempo e, conseqüentemente, intensificando o desgaste mecânico dos atuadores. Ao contrário da energia, que é limitada superiormente pela saturação, o índice TV não possui esse teto natural, crescendo indefinidamente com r .

Finalmente, verifica-se que, do ponto de vista energético, fatores de superamostragem baixos como $r = 4$ ou $r = 8$ já são suficientes para atingir o impacto máximo. No entanto, do ponto de vista do desgaste do atuador, aumentos adicionais em r amplificam progressivamente o dano imposto ao sistema físico, o que representa uma dimensão adicional de risco não capturada apenas pela métrica de energia.

4.1.5 Detecção do Ataque de Superamostragem

Para identificar a presença do ataque de superamostragem, emprega-se um detector de anomalias baseado no filtro de *Kalman* operando em uma taxa de amostragem superior à do sistema original.

Neste trabalho, o filtro de *Kalman* foi configurado para operar com período $T_{\text{KF}} = T_D/2$, ou seja, com taxa de amostragem duas vezes superior à do sistema original. Para os experimentos de detecção, o tempo total de simulação foi estendido para 45 s, sendo o ataque ativado aos 22,5 s. Essa extensão temporal foi necessária para permitir a aplicação adequada do método de detecção baseado na função de autocorrelação, que requer janelas de observação suficientemente longas para capturar a estrutura de correlação do erro. Com uma janela deslizante de tamanho $W = 190$ amostras, o período estendido garante dados suficientes tanto para a fase de operação normal quanto para a fase sob ataque, possibilitando uma análise estatística mais confiável.

As matrizes do filtro ($\bar{\mathbf{A}}$, $\bar{\mathbf{b}}$, $\bar{\mathbf{C}}$, $\bar{\mathbf{b}}_w$) foram obtidas por meio da discretização do modelo contínuo com período T_{KF} , mantendo-se constantes ao longo de toda a simulação. Os parâmetros de configuração do filtro de *Kalman* foram definidos por meio de testes de simulação, buscando o melhor compromisso entre precisão de estimação e rejeição de ruídos:

$$\mathbf{P}_0 = \mathbf{I}_2, \quad \bar{q} = 5, \quad \bar{\mathbf{R}} = \text{diag} \left(5 \times 10^{-6}, 0,01 \right), \quad (4.15)$$

em que $\mathbf{P}_0$ representa a matriz de covariância inicial do erro de estimação, $\bar{q}$ é a variância do ruído de processo no modelo do filtro, e $\bar{\mathbf{R}}$ é a matriz de covariância do ruído de

medição.

A detecção do ataque foi realizada por meio de dois métodos complementares: (i) análise da estatística qui-quadrado (χ^2) baseada no índice $g(j)$, e (ii) análise da função de autocorrelação (ACF) do erro de estimação. Ambos os métodos são descritos nas subseções seguintes para os problemas de regulação e rastreamento.

4.1.5.1 Detecção no Problema de Regulação

A Figura 17 apresenta os resultados da detecção do ataque de superamostragem para o problema de regulação. O painel superior mostra o índice de detecção $g(j)$ baseado na estatística qui-quadrado, enquanto o painel inferior exhibe o resultado da detecção por meio da análise da autocorrelação do erro.

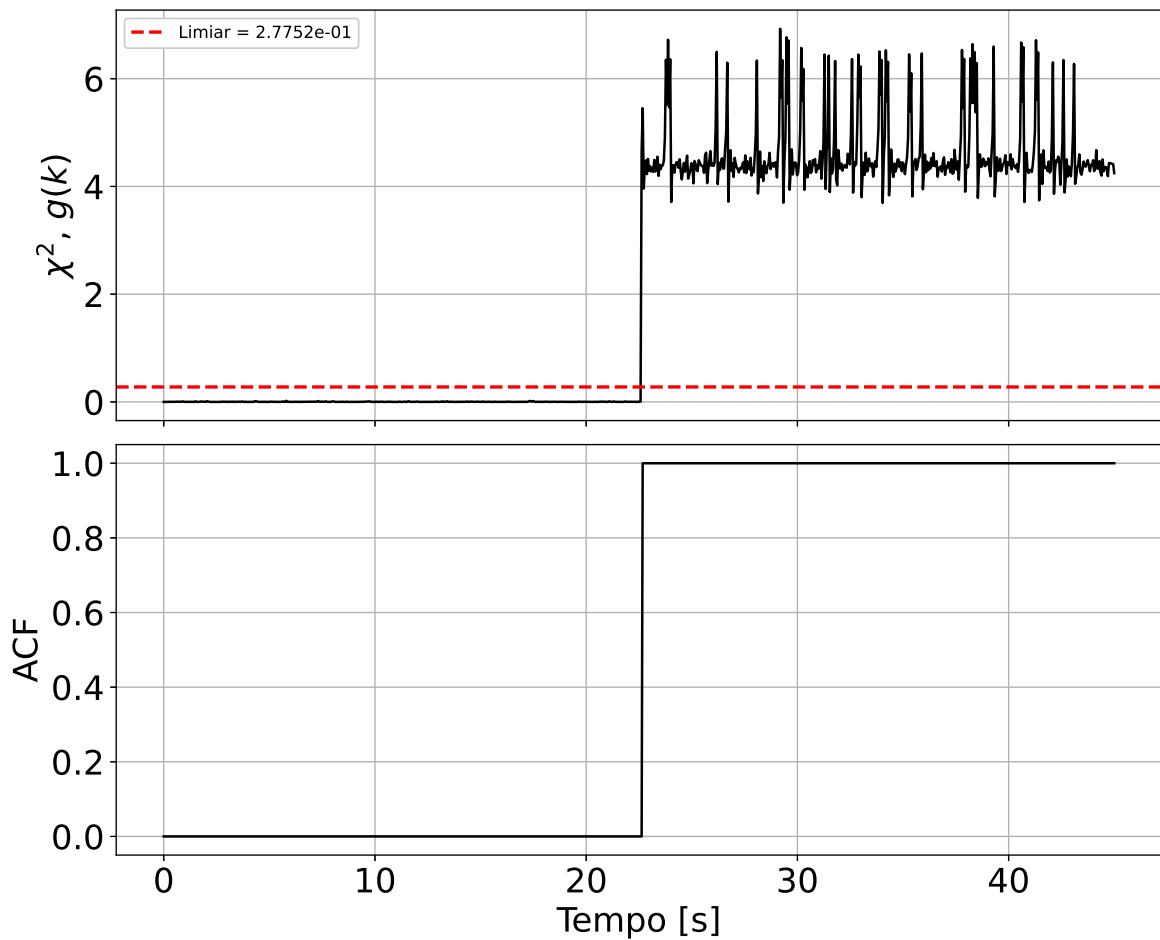


Figura 17 – Resultados da detecção do ataque de superamostragem para o problema de regulação. O painel superior mostra $g(j)$. O painel inferior exhibe a ACF.

Para o método baseado na estatística qui-quadrado, o limiar de detecção foi estabelecido utilizando a heurística descrita na Equação (3.28), resultando em $\Psi_{\chi^2} = 0,2775$.

Observa-se na Figura 17 (painel superior) que, antes da ativação do ataque (até 22,5 s), o índice $g(j)$ permanece consistentemente abaixo do limiar estabelecido, indicando operação normal do sistema. Após a ativação do ataque, há um aumento abrupto e sustentado no valor de $g(j)$, ultrapassando significativamente o limiar de detecção.

Para o método baseado na função de autocorrelação, o limiar foi determinado conforme a Equação (3.33), considerando uma janela deslizante de tamanho $W = 190$ amostras e um nível de confiança de 99,9% ($\alpha = 0,001$), resultando em $\Psi_{ACF} = 0,239$. Quando a ACF ultrapassa este limiar, o alarme é acionado, sendo indicado graficamente por 1 (detectado) ou 0 (não detectado).

Conforme ilustrado no painel inferior da Figura 17, a autocorrelação do erro permanece dentro dos limites esperados durante operação normal, consistente com a hipótese de erro branco. Após a ativação do ataque, observam-se violações do limiar, indicando que o erro deixou de apresentar características de processo branco. Essa mudança de comportamento é consequência direta da manipulação introduzida pelo ataque, que impõe correlações temporais não modeladas na dinâmica do erro de estimação.

Ambos os métodos de detecção demonstraram eficiência na identificação do ataque de superamostragem no problema de regulação, com o método qui-quadrado apresentando resposta imediata à presença do ataque e o método de autocorrelação fornecendo confirmação adicional utilizando a análise de propriedades estatísticas de segunda ordem do erro.

4.1.5.2 Detecção no Problema de Rastreamento

A Figura 18 apresenta os resultados da detecção do ataque de superamostragem para o problema de rastreamento, seguindo a mesma estrutura de análise aplicada ao problema de regulação.

Utilizando a mesma metodologia descrita na Equação (3.28), o limiar foi $\Psi_{\chi^2} = 0,1603$. Esse valor, inferior ao observado no problema de regulação, reflete diferenças nas características estatísticas do erro de estimação devido à dinâmica de rastreamento e ao esforço de controle.

Conforme demonstrado no painel superior da Figura 18, o índice $g(j)$ mantém-se consistentemente abaixo do limiar durante a fase de operação normal (até 22,5 s). Após a ativação do ataque, observa-se uma elevação significativa e persistente do índice, ultrapassando o limiar de detecção. Esse comportamento confirma a capacidade do método de identificar o ataque mesmo em cenários onde o sistema já opera com esforço de controle devido às exigências do rastreamento.

Para o método de autocorrelação, o limiar foi calculado conforme a Equação (3.33), utilizando os mesmos parâmetros e característica do alarme do problema de regulação

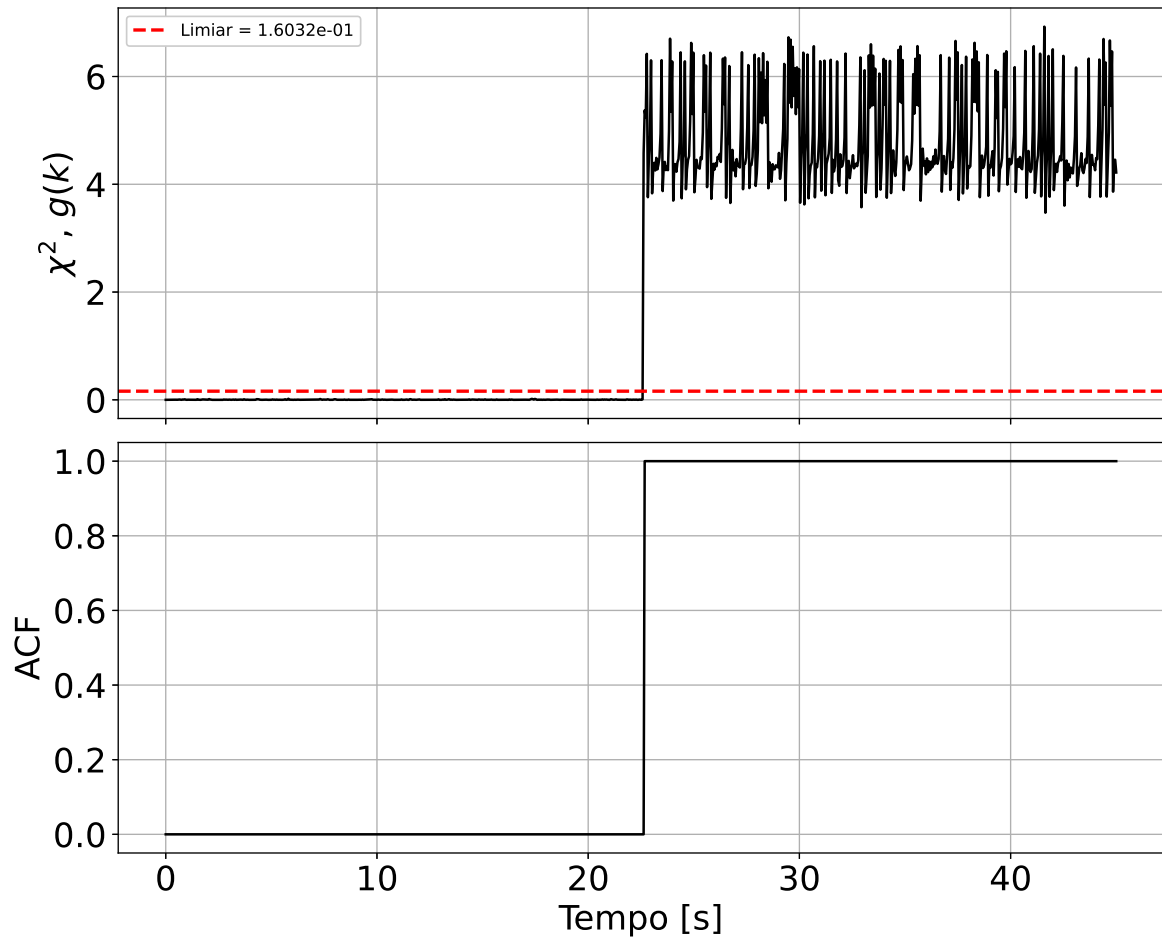


Figura 18 – Resultados da detecção do ataque de superamostragem para o problema de rastreamento. Para mais informações, ver legenda da Figura 17.

($W = 190$, $\alpha = 0,001$), resultando em $\Psi_{ACF} = 0,239$.

O painel inferior da Figura 18 evidencia que esse padrão, é consistente com o observado no problema de regulação.

A análise comparativa entre os resultados de regulação e rastreamento revela que, embora os limiares absolutos sejam diferentes devido às características específicas de cada problema, ambos os métodos de detecção apresentam desempenho satisfatório.

4.1.6 Extensão para Sistemas MIMO: Aeronave F-8C DFBW

Para validar a extensão do ataque de superamostragem a sistemas com múltiplas entradas, foi considerado o modelo da aeronave F-8C *Digital Fly-By-Wire* (DFBW) da NASA (do inglês *National Aeronautics and Space Administration*), um caça naval modificado para substituir o sistema de controle mecânico por um sistema eletrônico digital (ELLIOTT, 1977).

4.1.6.1 Descrição do Sistema

O modelo linearizado da aeronave F-8C representa a dinâmica lateral-direcional, que controla os movimentos de rolagem (*roll*) e guinada (*yaw*). O vetor de estados e o vetor de entradas são definidos como:

$$\mathbf{x} = \begin{bmatrix} p \\ r \\ \beta \\ \phi \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} \delta_a \\ \delta_r \end{bmatrix}, \quad (4.16)$$

em que p é a taxa de rolagem [rad/s], r é a taxa de guinada [rad/s], β é o ângulo de derrapagem [rad], ϕ é o ângulo de rolagem [rad], δ_a é a deflexão do *aileron* [rad] e δ_r é a deflexão do leme [rad].

As matrizes contínuas do modelo linearizado, extraídas de Elliott (1977) para a condição de voo a *Mach* 0,6 e altitude de 20.000 pés, são:

$$\mathbf{A}_c = \begin{bmatrix} -2,6 & 0,25 & -38,0 & 0 \\ -0,075 & -0,27 & 4,4 & 0 \\ 0,078 & -0,99 & -0,23 & 0,052 \\ 1,0 & 0,078 & 0 & 0 \end{bmatrix}, \quad \mathbf{B}_c = \begin{bmatrix} 17,0 & 7,0 \\ 0,82 & -3,2 \\ 0 & 0,046 \\ 0 & 0 \end{bmatrix}. \quad (4.17)$$

Esse modelo captura os acoplamentos cruzados característicos da dinâmica lateral-direcional de aeronaves, onde os *ailerons* afetam não apenas a rolagem mas também a guinada (devido ao momento adverso de guinada), e o leme afeta simultaneamente a guinada e a derrapagem lateral.

4.1.6.2 Configuração da Simulação

A Tabela 4 apresenta os parâmetros utilizados na simulação do ataque de superamostragem no sistema F-8C DFBW.

Os limites de deflexão das superfícies de controle foram obtidos das especificações do sistema de atuação do F-8C DFBW. O controlador nominal foi projetado via DLQR priorizando a redução do ângulo de derrapagem β e do ângulo de rolagem ϕ para garantir coordenação lateral adequada.

4.1.6.3 Evolução das Variáveis de Estado

A Figura 19 apresenta a evolução temporal das quatro variáveis de estado do sistema F-8C. Observa-se que as variáveis de estado do sistema atacado (em vermelho) coincidem com as variáveis de estado nominais (em azul) nos instantes de amostragem original $kT_D = 0,02$ s. Entre os instantes de amostragem, o sistema atacado apresenta

Tabela 4 – Parâmetros da simulação do ataque MIMO no sistema F-8C DFBW.

Parâmetro	Descrição	Valor	Unidade
n	Número de estados	4	–
m	Número de entradas	2	–
T_D	Período nominal (50 Hz)	0,02	s
r	Fator de superamostragem	4	–
T_d	Período do atacante (200 Hz)	0,005	s
t_{sim}	Tempo de simulação	5	s
$\mathbf{x}_0$	Condição inicial	$[0,05; 0,02; 0,03; 0,05]^T$	rad/s; rad/s; rad; rad
<i>Limites de saturação</i>			
$\bar{\delta}_a$	Limite superior <i>aileron</i>	0,436 (25°)	rad
$\underline{\delta}_a$	Limite inferior <i>aileron</i>	–0,436 (–25°)	rad
$\bar{\delta}_r$	Limite superior leme	0,524 (30°)	rad
$\underline{\delta}_r$	Limite inferior leme	–0,524 (–30°)	rad
<i>Ponderações do DLQR</i>			
Q	Matriz de ponderação dos estados	diag(1, 1, 50, 20)	–
R	Matriz de ponderação das entradas	$\mathbf{I}_2$	–

oscilações de alta frequência, particularmente pronunciadas na taxa de guinada r e na taxa de rolagem p .

4.1.6.4 Ações de Controle

A Figura 20 ilustra as deflexões das superfícies de controle aplicadas ao sistema. O controle nominal (em azul) apresenta um comportamento suave, com deflexões que convergem gradualmente para zero à medida que os estados se estabilizam. Em contraste, o controle de ataque (em vermelho) oscila de forma agressiva, frequentemente atingindo os limites físicos de deflexão: $\pm 25^\circ$ para o *aileron* e $\pm 30^\circ$ para o leme.

O comportamento *bang-bang* resultante é problemático em aeronaves reais, pois pode induzir modos aeroelásticos não modelados, excitar frequências estruturais ressonantes e acelerar significativamente o desgaste dos servomecanismos de atuação.

4.1.6.5 Métricas de Impacto

A Tabela 5 resume as métricas de impacto do ataque de superamostragem no sistema F-8C DFBW.

O sistema sob ataque apresentou $E_{\text{RMS}} = 0,6661$, enquanto o sistema nominal registrou $E_{\text{RMS}} = 0,0361$, resultando em uma razão de amplificação de $18,47\times$. Este aumento indica que o ataque de superamostragem injeta mais energia no sistema através das superfícies de controle, o que se traduz em maior consumo dos atuadores hidráulicos e maior demanda sobre o sistema de potência da aeronave.

A taxa de incremento energético $\Delta E = 1747,50\%$ representa o aumento percentual da energia consumida pelo sistema atacado em relação ao nominal. Este valor demonstra

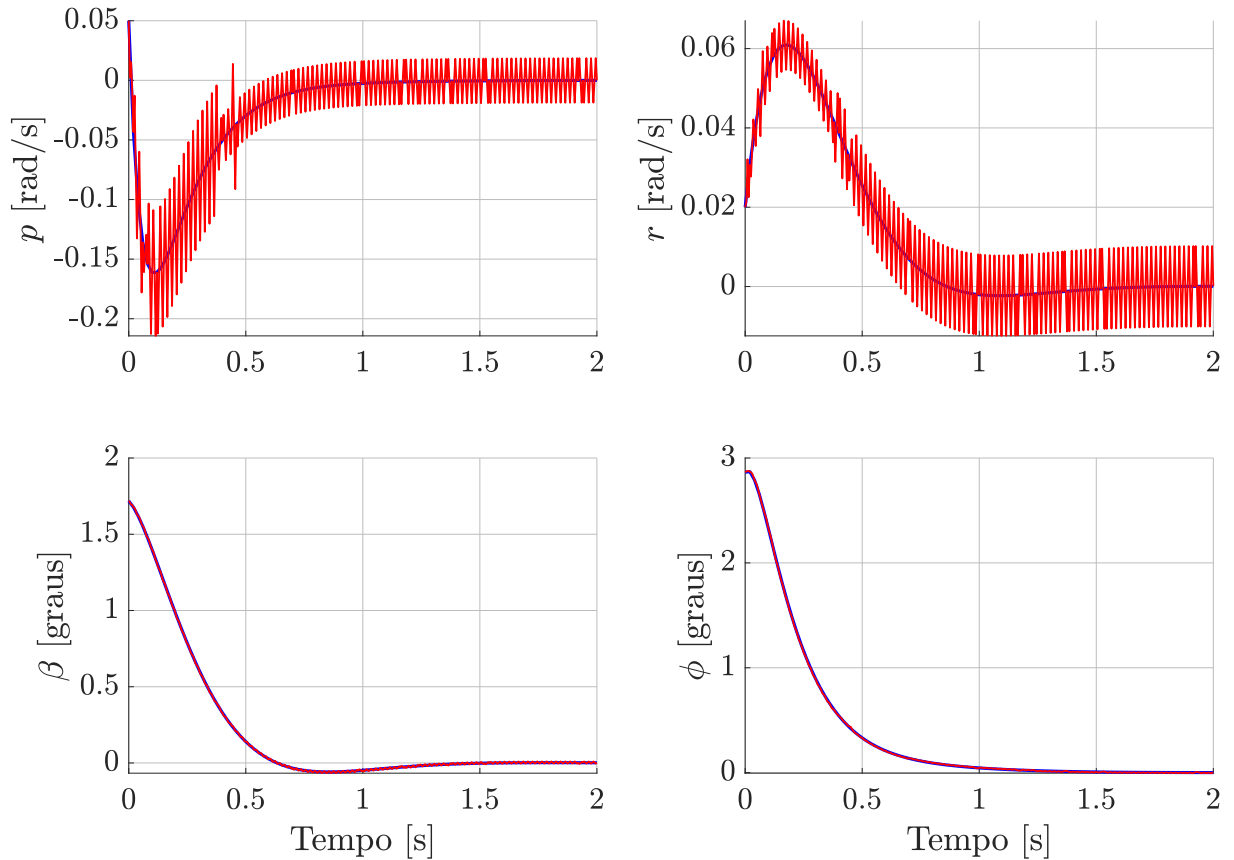


Figura 19 – Evolução temporal das variáveis de estado da aeronave F-8C DFBW. Em azul: sistema sob controle nominal com período $T_D = 0,02$ s (50 Hz). Em vermelho: sistema sob ataque de superamostragem com período $T_d = 0,005$ s (200 Hz). Paineil superior esquerdo: taxa de rolagem p . Paineil superior direito: taxa de guinada r . Paineil inferior esquerdo: ângulo de derrapagem β . Paineil inferior direito: ângulo de rolagem ϕ .

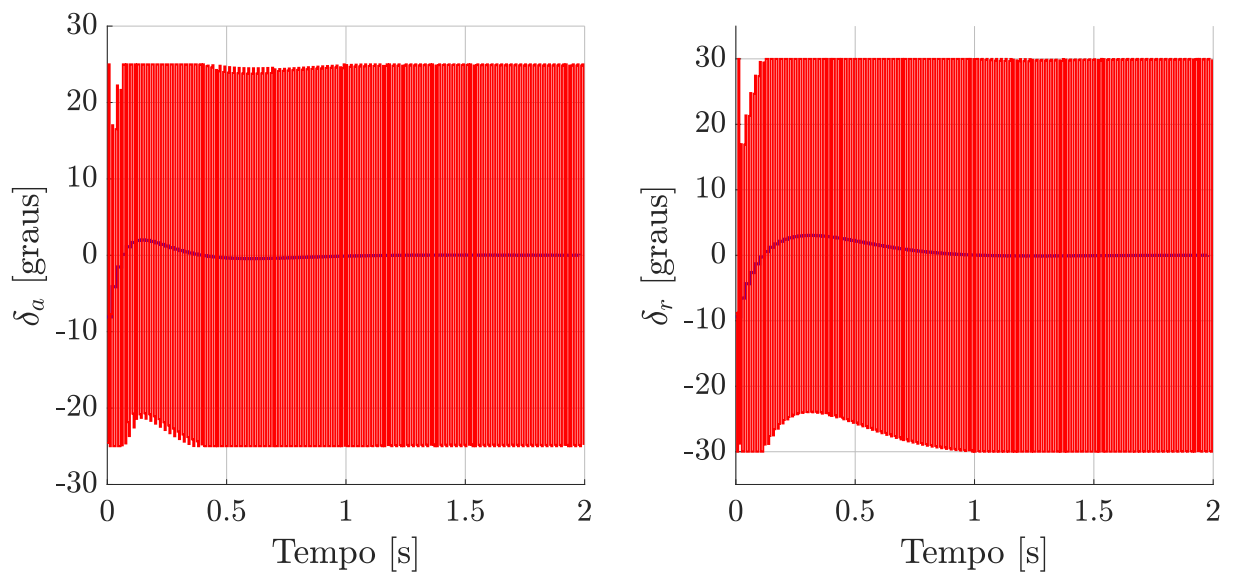


Figura 20 – Deflexões das superfícies de controle da aeronave F-8C DFBW. Em azul: controle nominal. Em vermelho: controle de ataque de superamostragem. Paineil esquerdo: deflexão do *aileron* δ_a (limite: $\pm 25^\circ$). Paineil direito: deflexão do leme δ_r (limite: $\pm 30^\circ$).

Tabela 5 – Métricas de impacto do ataque de superamostragem no sistema F-8C DFBW.

Métrica	Ataque	Nominal
<i>Energia RMS – Equação (3.14)</i>		
E_{RMS}	0,6661	0,0361
Razão de amplificação	18,47×	
<i>Taxa de Incremento Energético – Equação (3.15)</i>		
ΔE (%)	1747,50%	
<i>Índice de Desgaste do Atuador – Equação (3.16)</i>		
TV	412,65	0,50
Razão de desgaste	825,54×	

que, embora o ataque mantenha a condição de coincidência nos instantes de amostragem, a redistribuição temporal das ações de controle resulta em um consumo energético quase 18 vezes superior.

O índice de desgaste, medido pela variação total (TV), apresentou o resultado mais crítico: $TV = 412,65$ para o sistema atacado contra $TV = 0,50$ para o nominal, resultando em uma razão de desgaste de $825,54\times$. O aumento de mais de 800 vezes na variação total tem implicações para a manutenção e confiabilidade do sistema. Atuadores aeronáuticos são projetados para suportar um número finito de ciclos de operação antes de requererem inspeção ou substituição.

4.2 Cenário 2: Explorando o Ataque de Superamostragem em Sistemas Não Lineares

Nesse cenário, investiga-se a extensão do ataque de superamostragem para sistemas não lineares. Diferentemente do Cenário 1, onde o sistema era linear, aqui considera-se um motor de corrente contínua (DC, do inglês *direct current*) com saturação magnética, cuja dinâmica apresenta não linearidades que exigem técnicas específicas de modelagem e controle.

4.2.1 Descrição do Sistema e Configuração Experimental

Antes de formular o ataque, apresenta-se o modelo matemático do sistema-alvo, cujas não linearidades motivam o uso das abordagens descritas nas subseções seguintes.

4.2.1.1 Modelo do Motor DC com Saturação Magnética

O motor de corrente contínua é um dos atuadores mais utilizados em aplicações industriais, desde sistemas de posicionamento até acionamentos de velocidade variável.

O modelo clássico do motor DC assume uma relação linear entre corrente e torque, bem como entre velocidade e força contra-eletromotriz. No entanto, em condições de operação com correntes elevadas ou em motores de menor porte, a saturação magnética do núcleo ferromagnético introduz não linearidades que afetam o desempenho do sistema de controle (FITZGERALD; KINGSLEY; UMANS, 2003).

O modelo simplificado de segunda ordem do motor DC com saturação magnética é dado por:

$$\begin{aligned}\dot{x}_1 &= -\frac{R}{L}x_1 - \frac{K_e}{L}\tanh(x_2) + \frac{1}{L}u \\ \dot{x}_2 &= \frac{K_t}{J}\tanh(x_1) - \frac{B}{J}x_2\end{aligned}\tag{4.18}$$

em que x_1 [A] é a corrente de armadura, representando a corrente elétrica que flui pelo enrolamento de armadura do motor e é proporcional ao torque eletromagnético desenvolvido; x_2 [rad/s] é a velocidade angular do rotor, correspondendo à velocidade de rotação do eixo do motor; u [V] é a tensão de armadura, entrada de controle do sistema aplicada aos terminais do enrolamento de armadura; R [Ω] é a resistência elétrica do enrolamento de armadura; L [H] é a indutância do enrolamento de armadura; K_e [V·s/rad] é a constante de força contra-eletromotriz; K_t [N·m/A] é a constante de torque do motor; J [kg·m²] é o momento de inércia do rotor; e B [N·m·s/rad] é o coeficiente de atrito viscoso. A função $\tanh(\cdot)$ modela a saturação magnética, representando o efeito de que, para valores elevados de corrente ou velocidade, o fluxo magnético não cresce linearmente, mas satura devido às características do material ferromagnético do núcleo.

4.2.1.2 Parâmetros do Motor

Os parâmetros utilizados na simulação são valores típicos de um motor DC de pequeno porte, apresentados na Tabela 6.

Tabela 6 – Parâmetros do motor DC com saturação magnética.

Parâmetro	Descrição	Valor	Unidade
R	Resistência de armadura	1,0	Ω
L	Indutância de armadura	0,5	H
K_e	Constante de força contra-eletromotriz	0,01	V·s/rad
K_t	Constante de torque	0,01	N·m/A
J	Momento de inércia	0,01	kg·m ²
B	Coeficiente de atrito viscoso	0,1	N·m·s/rad

4.2.1.3 Dados da Simulação

A configuração da simulação do ataque de superamostragem é apresentada na Tabela 7.

Tabela 7 – Configuração da simulação do ataque no motor DC.

Parâmetro	Descrição	Valor	Unidade
T_D	Período de amostragem original	0,1	s
r	Fator de superamostragem	4	–
T_d	Período de amostragem do atacante	0,025	s
T_{sim}	Tempo total de simulação	3	s
$\mathbf{x}_0$	Condição inicial	$[2; 2]^T$	–
$\bar{u}$	Limite superior do atuador	24	V
$\underline{u}$	Limite inferior do atuador	–24	V

4.2.2 Formulação do Ataque via Modelagem *Takagi-Sugeno*

A formulação do ataque de superamostragem para o motor DC com saturação magnética segue a metodologia apresentada na Seção 3.3.1, utilizando a representação *Takagi-Sugeno* (T-S) via não linearidade de setor.

4.2.2.1 Representação T-S via Não Linearidade de Setor

O modelo descrito na Equação (4.18) pode ser reescrito na forma *quasi*-LPV identificando as não linearidades:

$$\begin{aligned} z_1(x_1) &= \frac{\tanh(x_1)}{x_1} \\ z_2(x_2) &= \frac{\tanh(x_2)}{x_2}. \end{aligned} \quad (4.19)$$

Para a região de operação considerada, os limites dos setores são:

$$\underline{z} = 0,1, \quad \bar{z} = 1,0. \quad (4.20)$$

O limite superior $\bar{z} = 1$ corresponde ao comportamento linear quando $x \rightarrow 0$ (em que $\tanh(x)/x \rightarrow 1$), enquanto o limite inferior $\underline{z} = 0,1$ representa a região de saturação para valores elevados de $|x|$.

4.2.2.2 Funções de Pertinência

As funções de pertinência para cada variável de premissa são construídas conforme a técnica de não linearidade de setor:

$$\begin{aligned} M_1^1(z_i) &= \frac{\bar{z} - z_i}{\bar{z} - \underline{z}} = \frac{1 - z_i}{0,9} \quad (\text{pertinência a } \underline{z}) \\ M_1^2(z_i) &= \frac{z_i - \underline{z}}{\bar{z} - \underline{z}} = \frac{z_i - 0,1}{0,9} \quad (\text{pertinência a } \bar{z}). \end{aligned} \quad (4.21)$$

Com duas variáveis de premissa (z_1 e z_2), obtém-se $r = 2^2 = 4$ regras *fuzzy*, cujas funções de pertinência normalizadas são:

$$\begin{aligned} h_1(\mathbf{z}) &= M_1^1(z_1) \cdot M_1^1(z_2) && \text{(Regra 1: } z_1 \rightarrow \underline{z}, z_2 \rightarrow \underline{z}) \\ h_2(\mathbf{z}) &= M_1^1(z_1) \cdot M_1^2(z_2) && \text{(Regra 2: } z_1 \rightarrow \underline{z}, z_2 \rightarrow \bar{z}) \\ h_3(\mathbf{z}) &= M_1^2(z_1) \cdot M_1^1(z_2) && \text{(Regra 3: } z_1 \rightarrow \bar{z}, z_2 \rightarrow \underline{z}) \\ h_4(\mathbf{z}) &= M_1^2(z_1) \cdot M_1^2(z_2) && \text{(Regra 4: } z_1 \rightarrow \bar{z}, z_2 \rightarrow \bar{z}). \end{aligned} \quad (4.22)$$

4.2.2.3 Matrizes dos Subistemas Lineares

As matrizes $\mathbf{A}_i$ para cada vértice do politopo são obtidas avaliando as matrizes do sistema nos extremos dos setores:

Vértice 1 ($z_1 = \underline{z}, z_2 = \underline{z}$):

$$\mathbf{A}_1 = \begin{bmatrix} -\frac{R}{L} & -\frac{K_e}{L}\underline{z} \\ \frac{K_t}{J}\underline{z} & -\frac{B}{J} \end{bmatrix} = \begin{bmatrix} -2 & -0,002 \\ 0,1 & -10 \end{bmatrix}. \quad (4.23)$$

Vértice 2 ($z_1 = \underline{z}, z_2 = \bar{z}$):

$$\mathbf{A}_2 = \begin{bmatrix} -\frac{R}{L} & -\frac{K_e}{L}\bar{z} \\ \frac{K_t}{J}\underline{z} & -\frac{B}{J} \end{bmatrix} = \begin{bmatrix} -2 & -0,02 \\ 0,1 & -10 \end{bmatrix}. \quad (4.24)$$

Vértice 3 ($z_1 = \bar{z}, z_2 = \underline{z}$):

$$\mathbf{A}_3 = \begin{bmatrix} -\frac{R}{L} & -\frac{K_e}{L}\underline{z} \\ \frac{K_t}{J}\bar{z} & -\frac{B}{J} \end{bmatrix} = \begin{bmatrix} -2 & -0,002 \\ 1 & -10 \end{bmatrix}. \quad (4.25)$$

Vértice 4 ($z_1 = \bar{z}, z_2 = \bar{z}$):

$$\mathbf{A}_4 = \begin{bmatrix} -\frac{R}{L} & -\frac{K_e}{L}\bar{z} \\ \frac{K_t}{J}\bar{z} & -\frac{B}{J} \end{bmatrix} = \begin{bmatrix} -2 & -0,02 \\ 1 & -10 \end{bmatrix}. \quad (4.26)$$

A matriz de entrada $\mathbf{B}$ é constante para todos os vértices:

$$\mathbf{B}_1 = \mathbf{B}_2 = \mathbf{B}_3 = \mathbf{B}_4 = \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}. \quad (4.27)$$

4.2.2.4 Discretização dos Modelos

Os subsistemas lineares foram discretizados utilizando o método ZOH com dois períodos de amostragem:

- **Modelo nominal:** discretizado com $T_D = 0,1$ s, obtendo as matrizes $\mathbf{A}_{D,i}$ e $\mathbf{B}_{D,i}$;
- **Modelo do atacante:** discretizado com $T_d = 0,025$ s, obtendo as matrizes $\mathbf{A}_{d,i}$ e $\mathbf{B}_{d,i}$;

4.2.2.5 Projeto do Controlador PDC

As matrizes de ponderação foram obtidas através de testes de simulação:

$$\mathbf{Q} = 10 \cdot \mathbf{I}_2, \quad R = 1. \quad (4.28)$$

Os ganhos obtidos para cada regra são apresentados na Tabela 8.

Tabela 8 – Ganhos do controlador DLQR para cada regra T-S.

Regra	$K_{i,1}$	$K_{i,2}$
$\mathbf{K}_1$	1,6448	0,0026
$\mathbf{K}_2$	1,6447	0,0001
$\mathbf{K}_3$	1,6515	0,0283
$\mathbf{K}_4$	1,6512	0,0257

A lei de controle nominal PDC é dada por:

$$u_k^* = - \sum_{i=1}^4 h_i(\mathbf{z}_k) \mathbf{K}_i \mathbf{x}_k. \quad (4.29)$$

4.2.2.6 Implementação do Ataque

O ataque de superamostragem foi implementado seguindo o Algoritmo 1, presente no Apêndice A, com a estrutura de otimização via espaço nulo. A cada instante de amostragem kT_D , o atacante:

1. Calcula as funções de pertinência $h_{i,k}$ com base no estado atual $\mathbf{x}_\ell$;
2. Obtém as matrizes agregadas $\mathbf{A}_d(h_k)$, $\mathbf{B}_d(h_k)$, $\mathbf{A}_D(h_k)$, $\mathbf{B}_D(h_k)$;
3. Constroi a matriz de controlabilidade $\mathcal{C}_r(h_k)$ e o vetor $\mathbf{d}_\ell(h_k)$;
4. Resolve o problema de otimização para maximizar a energia do ataque.

4.2.2.7 Métricas de Impacto do Ataque

Para quantificar o impacto do ataque de superamostragem no motor DC, foram avaliadas três métricas fundamentais: a energia RMS consumida pelo atuador, a taxa de incremento energético em relação ao controle nominal, e o índice de desgaste do atuador baseado na variação total do sinal de controle. Os resultados são apresentados na Tabela 9.

Tabela 9 – Métricas de impacto do ataque de superamostragem no motor DC via modelo T-S.

Métrica	Ataque	Nominal
<i>Energia RMS – Equação (3.14)</i>		
E_{RMS}	21,1124	0,7027
Razão de amplificação	30,05×	
<i>Taxa de Incremento Energético – Equação (3.15)</i>		
ΔE (%)	2904,63%	
<i>Índice de Desgaste do Atuador – Equação (3.16)</i>		
TV	3515,68	3,34
Razão de desgaste	1052,46×	

Os resultados demonstram que o ataque de superamostragem provoca um aumento de aproximadamente 30 vezes na energia RMS consumida pelo atuador, correspondendo a um incremento energético de 2904,63% em relação ao controle nominal. Além disso, o índice de desgaste do atuador, medido pela variação total do sinal de controle, é amplificado em mais de 1000 vezes.

4.2.2.8 Resultados e Discussão

A Figura 21 apresenta a evolução temporal das variáveis de estado do motor DC sob ataque de superamostragem.

A Figura 22 ilustra a ação de controle aplicada ao sistema.

4.2.3 Formulação do Ataque via Operadores de *Koopman*

Como alternativa à modelagem *Takagi-Sugeno*, essa seção apresenta a formulação do ataque de superamostragem utilizando operadores de *Koopman*, conforme a metodologia descrita na Seção 3.3.2. Essa abordagem baseada em dados permite capturar a dinâmica não linear do motor DC sem a necessidade de identificar explicitamente as funções de pertinência, utilizando uma representação linear em um espaço de dimensão elevada.

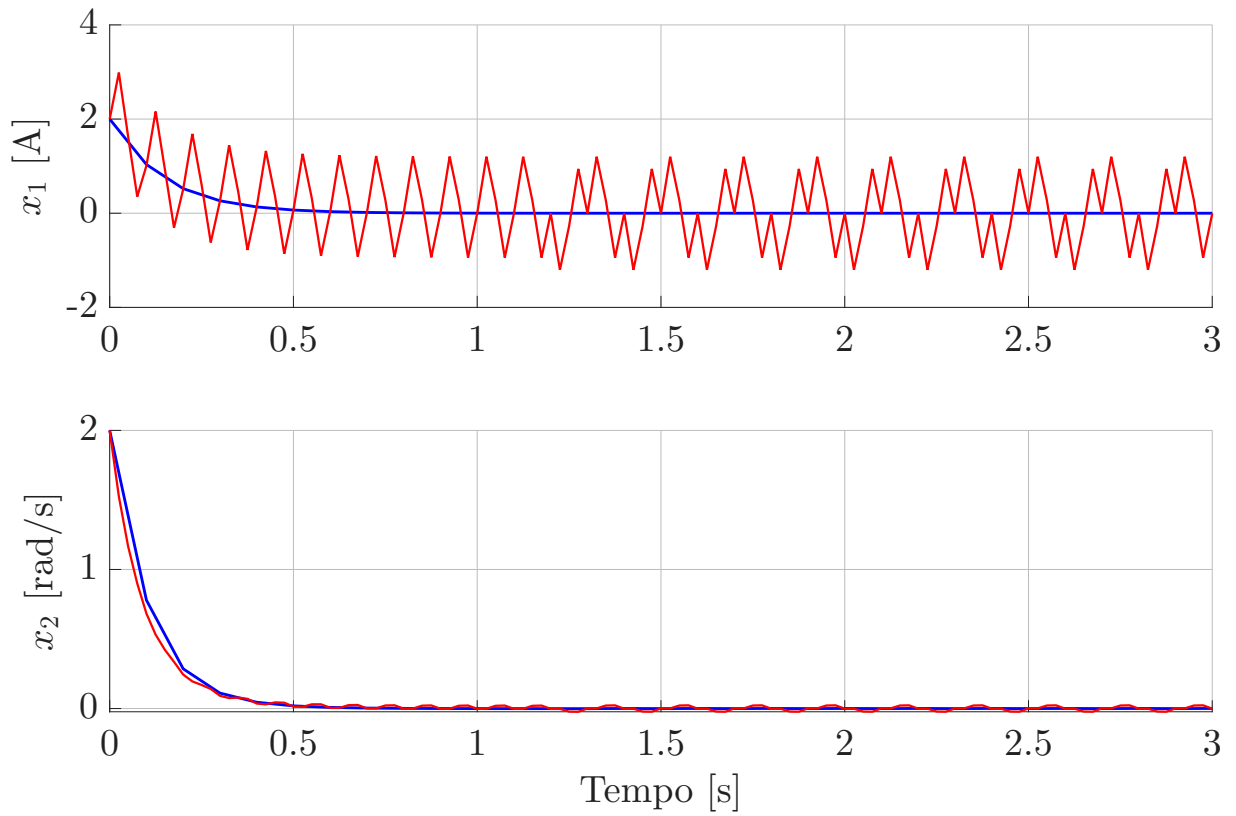


Figura 21 – Evolução temporal das variáveis de estado do motor DC. Em azul: sistema sob controle nominal com $T_D = 0,1$ s. Em vermelho: sistema sob ataque de superamostragem com $T_d = 0,025$ s.

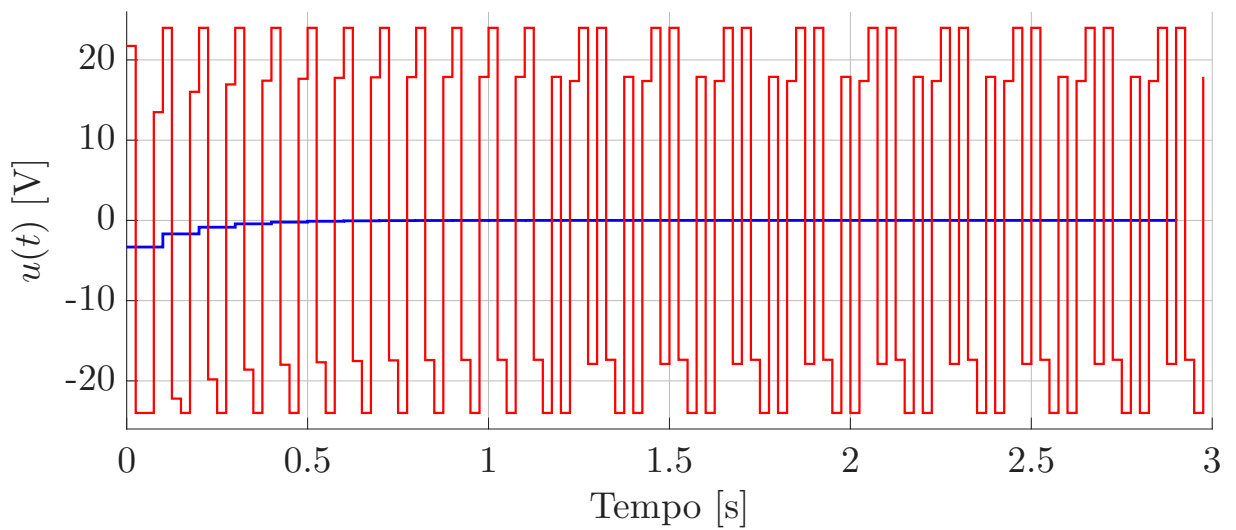


Figura 22 – Ação de controle aplicada ao motor DC. Em azul: controle nominal. Em vermelho: controle de ataque de superamostragem.

4.2.3.1 Identificação do Operador de Koopman via EDMD

A identificação do operador de *Koopman* foi realizada utilizando o método EDMD. Os dados de treinamento foram coletados por meio de simulações multi-episódio do sistema não linear apresentado na Equação (4.18), com os parâmetros apresentados na Tabela 10.

Tabela 10 – Configuração da identificação EDMD para o motor DC.

Parâmetro	Descrição	Valor	Unidade
$n_{\text{episódios}}$	Número de trajetórias de treinamento	200	–
$L_{\text{episódio}}$	Duração de cada trajetória	50	passos
N_{train}	Total de amostras de treinamento	10.000	–
$\mathbf{x}_0$	Condições iniciais aleatórias	$\mathcal{U}(-3, 3)^2$	A; rad/s
u	Entradas de excitação aleatórias	$\mathcal{U}(-24, 24)$	V
λ	Regularização de <i>Tikhonov</i>	10^{-4}	–

4.2.3.2 Escolha do Dicionário de Observáveis

A qualidade da aproximação do operador de *Koopman* depende fundamentalmente da escolha do dicionário de funções observáveis $\psi(\mathbf{x})$. Nesse trabalho, foram avaliadas duas configurações de dicionário com diferentes níveis de complexidade:

Biblioteca “Pobre” (Polinomial-Trigonométrica):

Essa configuração utiliza funções polinomiais, trigonométricas e a função $\tanh(\cdot)$, descrito por:

$$\psi_{\text{pobre}}(\mathbf{x}) = \begin{bmatrix} x_1 & x_1 \tanh(x_2) & \sin(2x_1) \\ x_2 & x_2 \tanh(x_1) & \cos(2x_1) \\ \tanh(x_1) & \sin(x_1) & \sin(2x_2) \\ \tanh(x_2) & \cos(x_1) & \cos(2x_2) \\ x_1 x_2 & \sin(x_2) & \sin(x_1) \cos(x_2) \\ x_1^2 + x_2^2 & \cos(x_2) & \sin(x_2) \cos(x_1) \\ x_1^2 & \sin(x_1 + x_2) & x_1 \sin(x_2) \\ x_2^2 & \cos(x_1 + x_2) & x_2 \cos(x_1) \\ x_1^3 & \sin(x_1 - x_2) & \tanh(x_1) \tanh(x_2) \\ x_2^3 & \cos(x_1 - x_2) & 1 \end{bmatrix} \quad (4.30)$$

A inclusão da função $\tanh(\cdot)$ no dicionário é motivada pelo conhecimento físico prévio do sistema: como a não linearidade do motor DC é a saturação magnética modelada por $\tanh(\cdot)$, sua presença explícita no dicionário permite que o EDMD capture diretamente essa característica sem necessidade de aproximá-la por outras funções.

Biblioteca “Rica” (Híbrida com RBF):

Essa configuração combina os estados lineares, a função de saturação conhecida e uma grade de funções de base radial (RBF) para capturar comportamentos não modelados:

$$\boldsymbol{\psi}_{\text{rica}}(\mathbf{x}) = \begin{bmatrix} x_1 \\ x_2 \\ \tanh(x_1) \\ \tanh(x_2) \\ \exp\left(-\frac{\|\mathbf{x}-\mathbf{c}_1\|^2}{2\sigma^2}\right) \\ \vdots \\ \exp\left(-\frac{\|\mathbf{x}-\mathbf{c}_{n_{\text{RBF}}}\|^2}{2\sigma^2}\right) \\ 1 \end{bmatrix}, \quad (4.31)$$

em que os centros $\mathbf{c}_i$ são distribuídos em uma grade uniforme no domínio $[-3, 3]^2$ com espaçamento de 1,5 unidades, resultando em $n_{\text{RBF}} = 25$ funções RBF com largura $\sigma = 1,5$.

4.2.3.3 Estrutura das Matrizes de *Koopman*

Após a identificação via EDMD, obtém-se a matriz de *Koopman* $\mathbf{K} = [\mathbf{A}_K \ \mathbf{B}_K]$ que relaciona os observáveis no instante k com os observáveis no instante $k + 1$:

$$\boldsymbol{\psi}(\mathbf{x}_{k+1}) \approx \mathbf{A}_K \boldsymbol{\psi}(\mathbf{x}_k) + \mathbf{B}_K u_k, \quad (4.32)$$

em que $\mathbf{A}_K \in \mathbb{R}^{N \times N}$ é a matriz de dinâmica de *Koopman* e $\mathbf{B}_K \in \mathbb{R}^{N \times 1}$ é o vetor de entrada no espaço de observáveis elevado.

Para o ataque de superamostragem, as matrizes são reamostradas para o período $T_d = T_D/r$ utilizando:

$$\bar{\mathbf{A}}_{d,K} = \exp(\mathbf{A}_{K,\text{cont}} T_d), \quad \bar{\mathbf{B}}_{d,K} = \mathbf{A}_{K,\text{cont}}^{-1} (\bar{\mathbf{A}}_K - \mathbf{I}) \mathbf{B}_{K,\text{cont}}, \quad (4.33)$$

em que $\bar{\mathbf{A}}_{d,K} \in \mathbb{R}^{N \times N}$ e $\bar{\mathbf{B}}_{d,K} \in \mathbb{R}^{N \times 1}$ são as matrizes discretizadas do sistema superamostrado.

A matriz de projeção física $\mathbf{C}_{\text{phys}} \in \mathbb{R}^{n \times N}$ extrai os estados físicos do espaço elevado:

$$\mathbf{C}_{\text{phys}} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n \times N}. \quad (4.34)$$

A matriz de controlabilidade projetada para o ataque é então:

$$\mathbf{C}_r^K = \mathbf{C}_{\text{phys}} \bar{\mathbf{\Gamma}}_K, \quad \text{em que} \quad \bar{\mathbf{\Gamma}}_K = \begin{bmatrix} \bar{\mathbf{A}}_{d,K}^{r-1} \bar{\mathbf{B}}_{d,K} & \bar{\mathbf{A}}_{d,K}^{r-2} \bar{\mathbf{B}}_{d,K} & \cdots & \bar{\mathbf{B}}_{d,K} \end{bmatrix} \in \mathbb{R}^{N \times r}, \quad (4.35)$$

em que $\mathbf{C}_r^K \in \mathbb{R}^{n \times r}$ é a matriz de controlabilidade do ataque no espaço físico.

O vetor de desvio $\mathbf{d}_k^K \in \mathbb{R}^n$, que representa a diferença entre o estado nominal futuro e a evolução livre do sistema superamostrado, é calculado por:

$$\mathbf{d}_k^K = \mathbf{C}_{\text{phys}} \left(\mathbf{z}_{k+1}^{\text{nom}} - \bar{\Theta}_K \mathbf{z}_k \right), \quad (4.36)$$

em que $\mathbf{z}_k = \boldsymbol{\psi}(\mathbf{x}_k) \in \mathbb{R}^N$ é o estado elevado no espaço de *Koopman*, $\mathbf{z}_{k+1}^{\text{nom}} = \mathbf{A}_K \mathbf{z}_k + \mathbf{B}_K \mathbf{u}_k^* \in \mathbb{R}^N$ é o estado nominal futuro no espaço de *Koopman*, $\bar{\Theta}_K = \bar{\mathbf{A}}_{d,K}^r \in \mathbb{R}^{N \times N}$ é a matriz de transição livre do sistema superamostrado, e $\mathbf{u}_k^* \in \mathbb{R}$ é a ação de controle nominal.

4.2.3.4 Controlador Nominal

Para simplificar, o controlador nominal foi projetado utilizando a linearização do sistema na origem. Para o motor DC com saturação magnética, descrito na Equação (4.18), a função $\tanh(\cdot)$ pode ser aproximada por $\tanh(x) \approx x$ quando $|x| \ll 1$, resultando no modelo linearizado:

$$\dot{\mathbf{x}} = \mathbf{A}_{\text{lin}} \mathbf{x} + \mathbf{b}_{\text{lin}} u, \quad (4.37)$$

com as matrizes:

$$\mathbf{A}_{\text{lin}} = \begin{bmatrix} -R/L & -K_e/L \\ K_t/J & -B/J \end{bmatrix} = \begin{bmatrix} -2 & -0,02 \\ 1 & -10 \end{bmatrix} \in \mathbb{R}^{2 \times 2}, \quad \mathbf{b}_{\text{lin}} = \begin{bmatrix} 1/L \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \in \mathbb{R}^{2 \times 1}. \quad (4.38)$$

O ganho DLQR foi calculado com as mesmas matrizes de ponderação utilizadas na formulação T-S, resultando no ganho de realimentação:

$$\mathbf{K}_{\text{nom}} = \text{dlqr}(\mathbf{A}_{D,\text{lin}}, \mathbf{B}_{D,\text{lin}}, \mathbf{Q}, R) \in \mathbb{R}^{1 \times 2}, \quad (4.39)$$

em que $\mathbf{A}_{D,\text{lin}} \in \mathbb{R}^{2 \times 2}$ e $\mathbf{B}_{D,\text{lin}} \in \mathbb{R}^{2 \times 1}$ são as matrizes do sistema linearizado discretizadas com período $T_D = 0,1$ s. Este ganho é aplicado em toda a região de operação, sendo uma aproximação válida nas proximidades da origem.

A lei de controle nominal é dada por:

$$u_k^* = -\mathbf{K}_{\text{nom}} \mathbf{x}_k \in \mathbb{R}. \quad (4.40)$$

4.2.3.5 Implementação do Ataque

O ataque de superamostragem via *Koopman* segue a mesma estrutura algorítmica da formulação T-S, com a diferença de que as matrizes $\mathbf{C}_r$ e $\mathbf{d}_\ell$ são calculadas a partir do modelo de *Koopman* identificado. A otimização foi realizada utilizando o método GRASP conforme o Algoritmo 2, presente no Apêndice A.

Tabela 11 – Comparação das métricas de impacto do ataque via *Koopman* para diferentes bibliotecas.

Métrica	Biblioteca “Pobre”	Biblioteca “Rica”	Nominal
<i>Energia RMS – Equação (3.14)</i>			
E_{RMS} [V]	17,64	19,73	0,85 / 0,71
Razão de amplificação	20,75×	27,89×	—
<i>Taxa de Incremento Energético – Equação (3.15)</i>			
ΔE (%)	1975,34	2688,89	—
<i>Índice de Desgaste do Atuador – Equação (3.16)</i>			
TV	2895,29	2948,98	3,36
Razão de desgaste	861,87×	878,90×	—

4.2.3.6 Comparação entre Bibliotecas

A Tabela 11 apresenta a comparação das métricas de impacto do ataque para as duas configurações de dicionário.

Os resultados demonstram que a biblioteca “rica” (com RBF) proporciona um ataque mais efetivo, com energia RMS 11,8% superior e taxa de incremento energético 36,1% maior em comparação à biblioteca “pobre”. Essa diferença deve-se à melhor capacidade das funções RBF em capturar a dinâmica não linear do sistema em toda a região de operação, resultando em previsões mais precisas do estado futuro e, consequentemente, em uma condição de coincidência melhor satisfeita.

4.2.3.7 Resultados

As Figuras 23 e 24 apresentam a evolução temporal das variáveis de estado do motor DC sob ataque de superamostragem para cada configuração de biblioteca.

As Figuras 25 e 26 ilustram a ação de controle aplicada ao sistema para cada biblioteca.

4.2.3.8 Análise do Erro de Coincidência

As Figuras 27 e 28 apresentam o erro de coincidência $\|\mathbf{x}_k^{\text{ataque}} - \mathbf{x}_k^{\text{nominal}}\|$ ao longo da simulação. Esse erro quantifica a diferença entre o estado do sistema atacado e o estado nominal nos instantes de amostragem original kT_D .

Observa-se que o erro de coincidência é maior nos instantes iniciais da simulação para ambas as bibliotecas. Esse comportamento é explicado por dois fatores:

1. **Região de operação:** A condição inicial $\mathbf{x}_0 = [2; 2]^T$ está distante da origem, região onde as não linearidades do sistema são mais evidentes. As funções polinomiais e

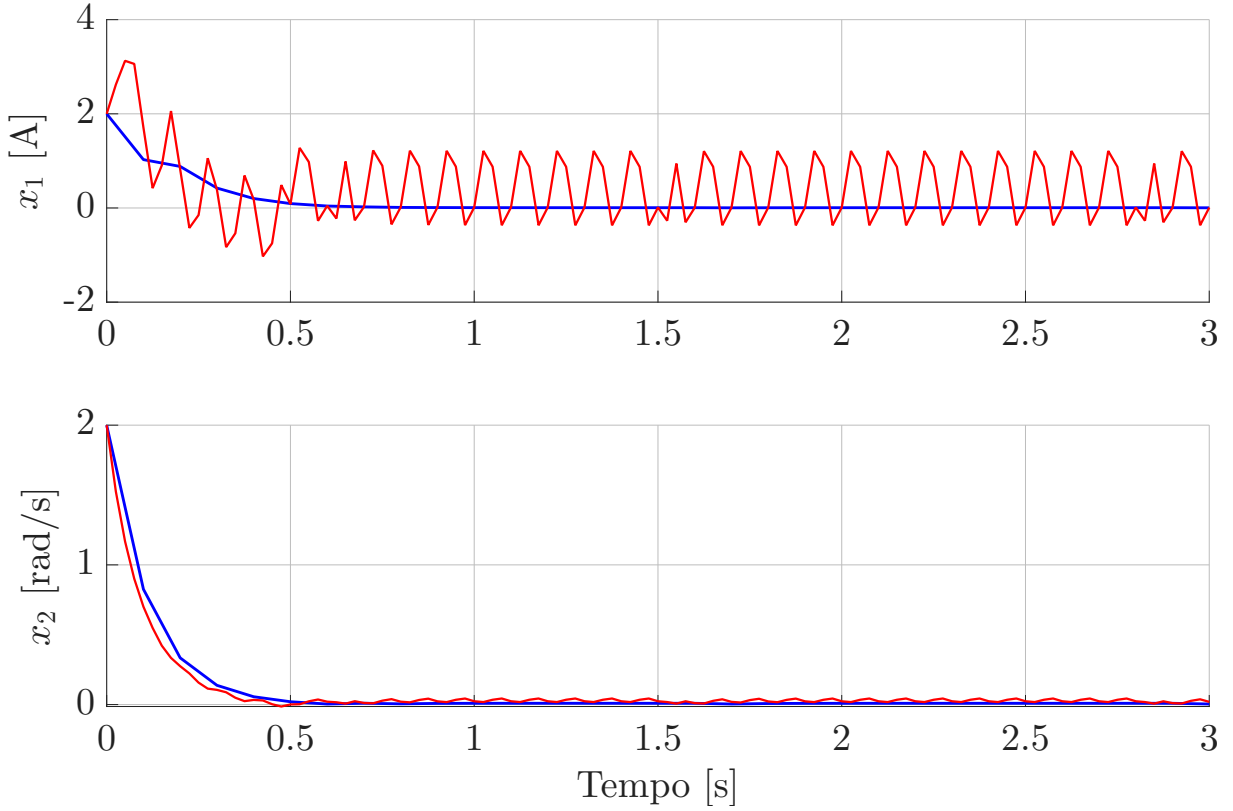


Figura 23 – Evolução temporal das variáveis de estado do motor DC utilizando operador de *Koopman* com biblioteca polinomial-trigonométrica. Para mais informações, ver legenda da Figura 21.

trigonométricas da biblioteca “pobre” têm maior dificuldade em aproximar a dinâmica nessa região, enquanto as RBF da biblioteca “rica”, distribuídas uniformemente no domínio, proporcionam melhor cobertura.

2. **Magnitude das ações de controle:** Nos instantes iniciais, tanto o controle nominal quanto o controle de ataque possuem magnitudes relevantes para estabilizar o sistema. Pequenos erros de predição do modelo de *Koopman* são amplificados por essas ações de controle, resultando em maior diferença entre os estados.

À medida que o sistema se aproxima da origem, onde a dinâmica é aproximadamente linear e bem capturada por ambas as bibliotecas, o erro de coincidência diminui.

4.3 Cenário 3: Perspectiva do Atacante - Reconhecimento, Preparação e Execução do Ataque

Enquanto o Cenário 1 apresentou a validação computacional do ataque de superamostragem em um ambiente simulado linear e o Cenário 2 estendeu essa validação para sistemas não lineares simulados, o presente cenário adota a perspectiva do atacante para demonstrar os passos necessários à implementação prática do ataque em um sistema físico

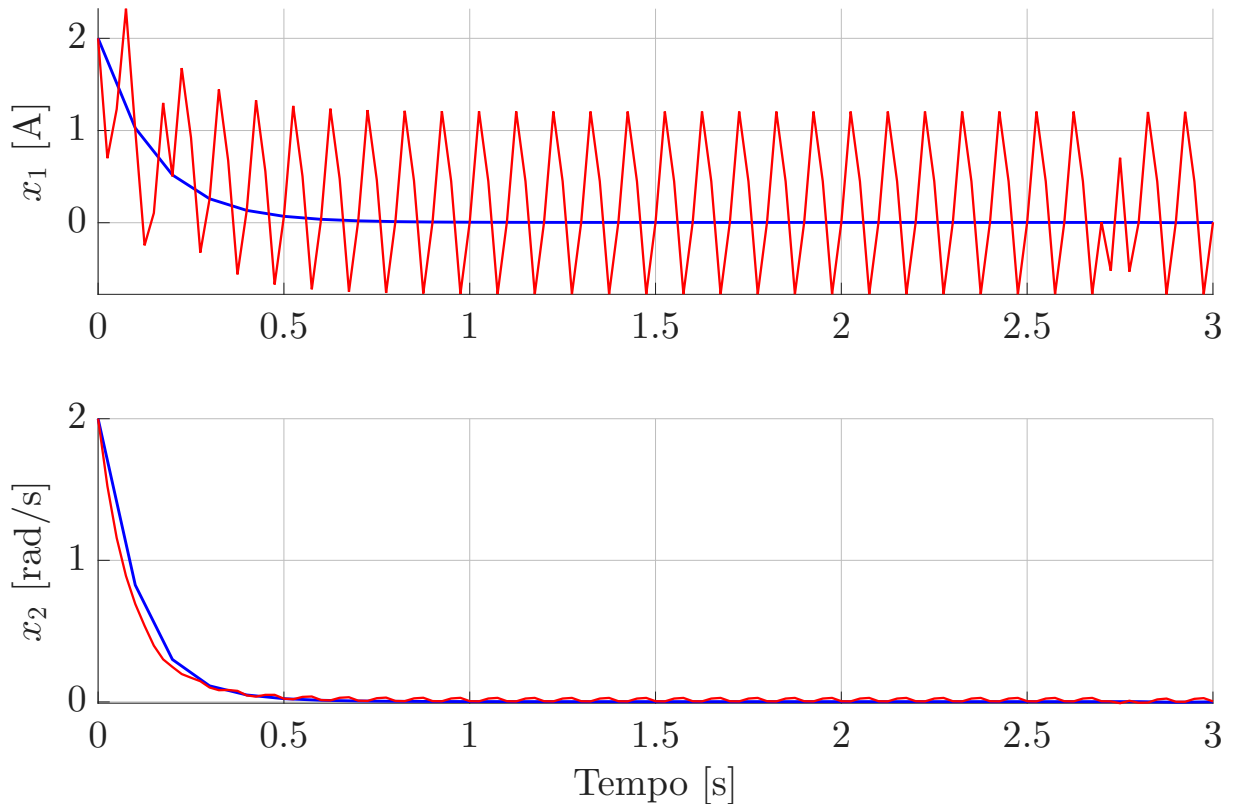


Figura 24 – Evolução temporal das variáveis de estado do motor DC utilizando operador de *Koopman* com biblioteca híbrida (RBF). Para mais informações, ver legenda da Figura 21.

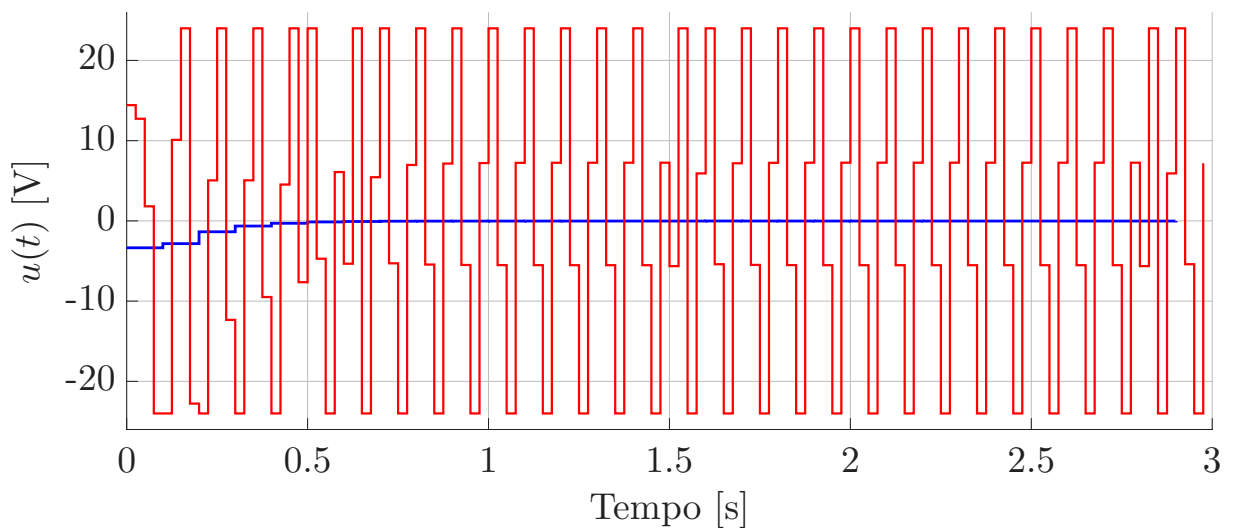


Figura 25 – Ação de controle aplicada ao motor DC utilizando operador de *Koopman* com biblioteca polinomial-trigonométrica. Para mais informações, ver legenda da Figura 22.

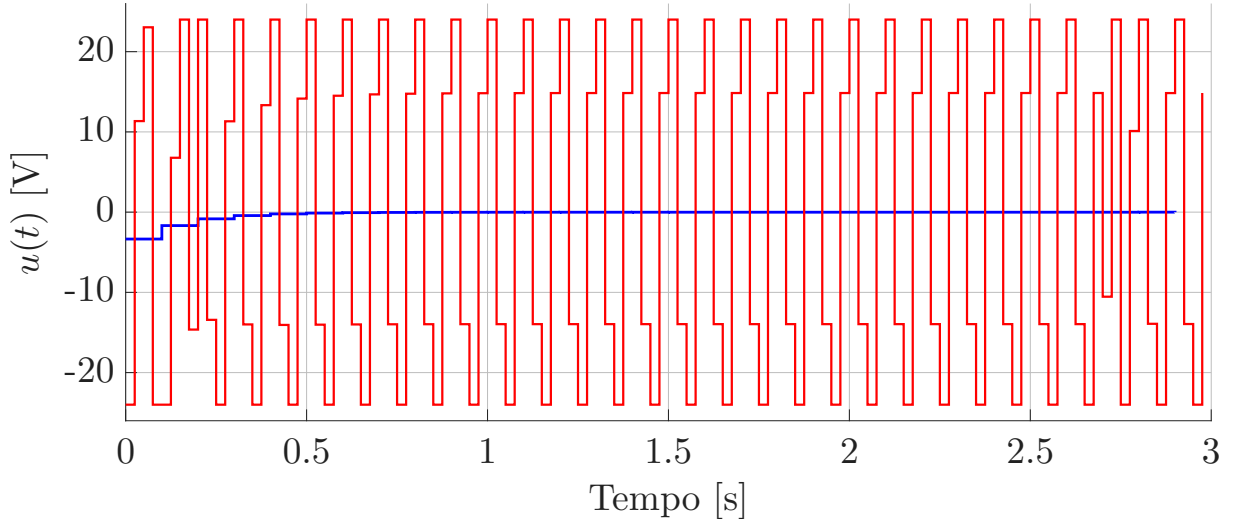


Figura 26 – Ação de controle aplicada ao motor DC utilizando operador de *Koopman* com biblioteca híbrida (RBF). Para mais informações, ver legenda da Figura 22.

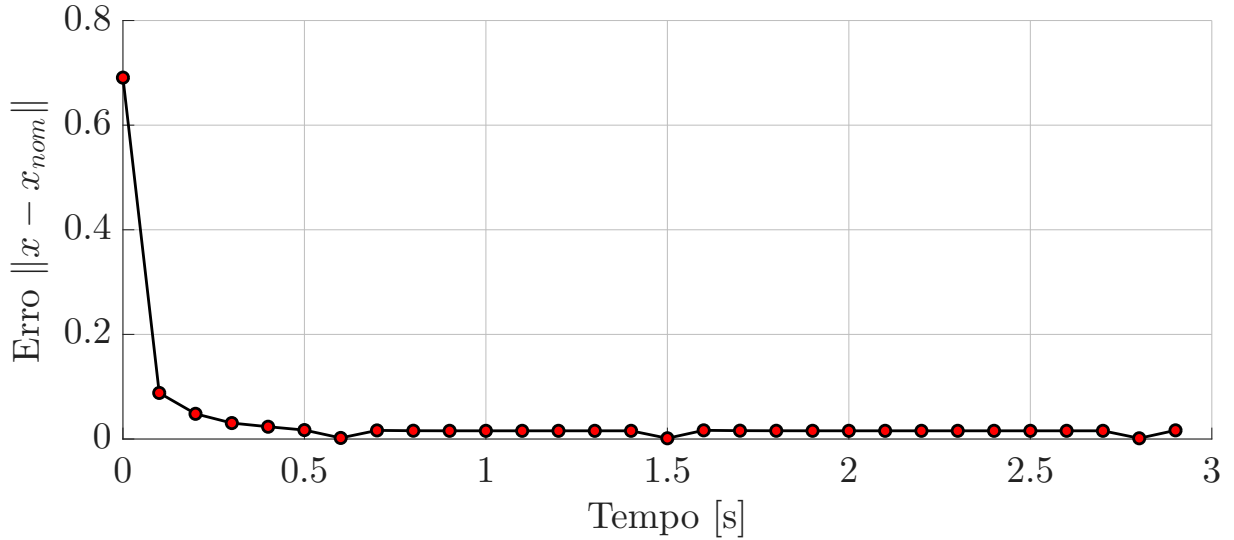


Figura 27 – Erro de coincidência $\|\mathbf{x}_k^{\text{ataque}} - \mathbf{x}_k^{\text{nominal}}\|$ utilizando biblioteca polinomial-trigonométrica.

real. Essa abordagem visa evidenciar as vulnerabilidades existentes em sistemas ciberfísicos convencionais e os procedimentos que um adversário capacitado poderia empregar para comprometer um sistema de controle.

O sistema-alvo escolhido para esse cenário é a planta piloto de controle de nível LAB-Volt 3503-M0, um sistema didático utilizado em laboratórios de ensino e pesquisa em controle de processos.

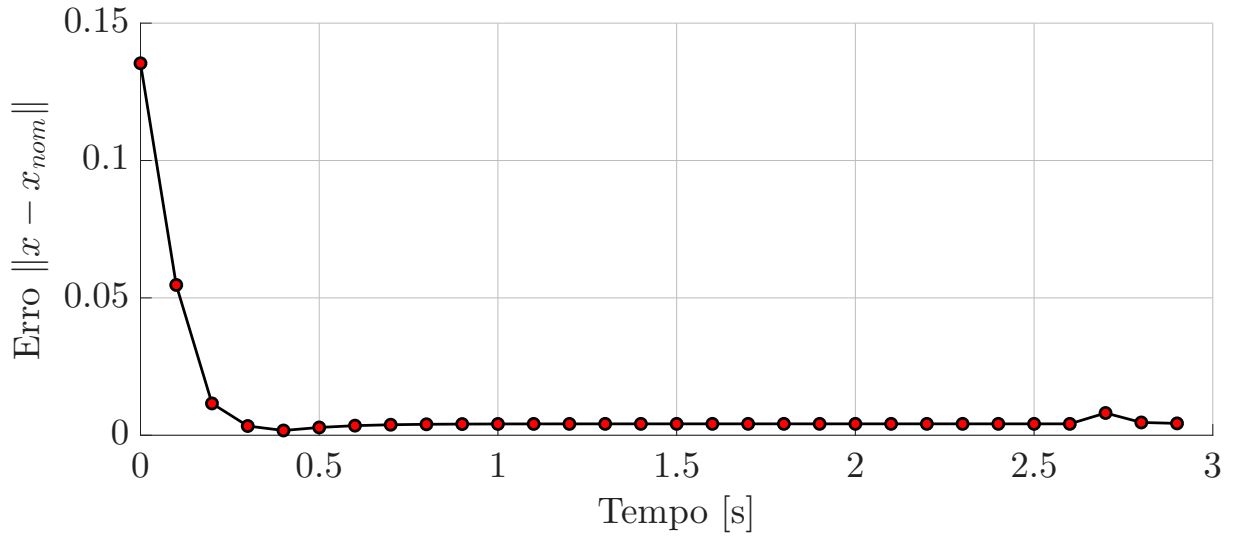


Figura 28 – Erro de coincidência $\|\mathbf{x}_k^{\text{ataque}} - \mathbf{x}_k^{\text{nominal}}\|$ utilizando biblioteca híbrida (RBF).

4.3.1 Reconhecimento do Sistema-Alvo

4.3.1.1 Planta Piloto LAB-Volt 3503-M0

A planta piloto LAB-Volt 3503-M0, é um sistema modular de controle de nível líquido desenvolvido pela Festo Didactic para demonstração e experimentação em controle de processos (FESTO DIDACTIC, 2015). O sistema foi projetado para treinamento prático em medição, controle e diagnóstico de falhas em processos de nível, apresentando tempos de resposta e dinâmicas representativas de processos industriais reais.

A Figura 29 apresenta o diagrama de processo e instrumentação (P&ID, do inglês *piping and instrumentation diagram*) da planta LAB-Volt 3503-M0, ilustrando a configuração dos componentes e as conexões hidráulicas e pneumáticas do sistema.

4.3.1.2 Sistema de Controle e Arquitetura Computacional

O sistema de controle implementado para a planta LAB-Volt 3503-M0 utiliza um microcontrolador ESP32 como plataforma de aquisição de dados e atuação. A arquitetura de controle adota uma topologia distribuída, onde o algoritmo de controle é executado em ambiente MATLAB® no computador hospedeiro, enquanto o ESP32 atua como interface de *hardware* entre o ambiente computacional e a planta física. A comunicação entre MATLAB® e ESP32 ocorre via protocolo serial (UART, do inglês *universal asynchronous receiver-transmitter*) por meio de conexão USB.

Essa arquitetura, embora funcional e adequada para propósitos didáticos, apresenta vulnerabilidades do ponto de vista de segurança ciberfísica. A comunicação serial não é criptografada, não possui mecanismos de autenticação, e o acesso físico ao computador hospedeiro ou aos cabos de comunicação permite interceptação e manipulação dos dados trocados entre os componentes do sistema.

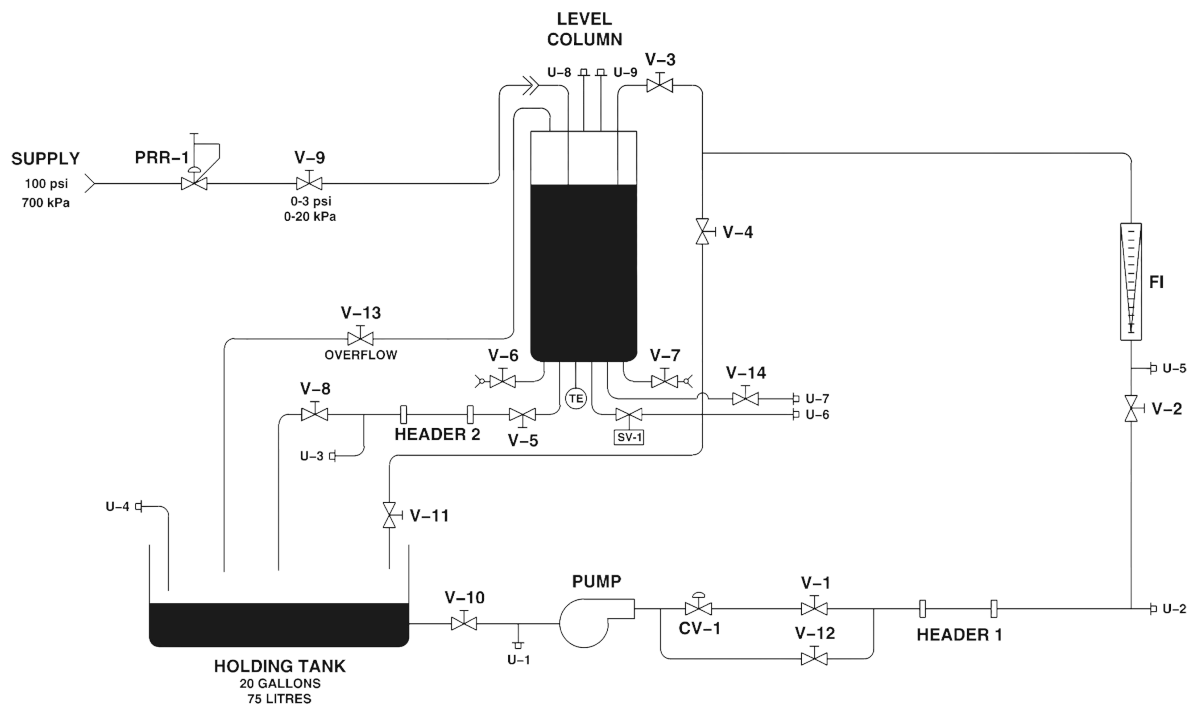


Figura 29 – Diagrama de processo e instrumentação (P&ID) da planta piloto LAB-Volt 3503-M0 (*Level Process Station*). O sistema inclui coluna de nível com tubo *bubbler*, válvula de controle pneumática (CV-1), bomba centrífuga, reservatório inferior (*holding tank*) de 75 L, *headers* de distribuição, rotâmetro (FI) e válvulas manuais (V-1 a V-14) para configuração do processo.

Fonte: FESTO DIDACTIC (2015).

4.3.1.3 Vetor de Ataque e Obtenção de Acesso

Para este cenário, assume-se que o atacante obteve acesso inicial ao sistema mediante técnicas de engenharia social combinadas com exploração de vulnerabilidades de segurança física. Especificamente, considera-se o seguinte vetor de ataque:

- **Fase 1: Reconhecimento preliminar:** O atacante obteve acesso físico ao laboratório onde o sistema está instalado. Durante visitas, o atacante observou a configuração do sistema, identificou os componentes utilizados (planta LAB-Volt, microcontrolador ESP32, *software* MATLAB®), e documentou a topologia da rede e os procedimentos operacionais;
- **Fase 2: Engenharia social:** O atacante obteve informações sobre os protocolos de comunicação utilizados, parâmetros de operação do sistema, e credenciais de acesso ao computador hospedeiro. Técnicas como *shoulder surfing* (observação discreta da digitação de senhas) e pretexto (solicitação de informações sob falsa justificativa) foram empregadas;
- **Fase 3: Instalação de *backdoor*:** Aproveitando-se de um momento em que o

sistema estava desacompanhado, o atacante conectou um dispositivo USB ao computador hospedeiro, executando um *script* que instalou um *backdoor* com privilégios elevados. Esse *backdoor* permite acesso remoto ao sistema e execução de comandos com privilégios administrativos;

- **Fase 4: Acesso remoto persistente:** Por meio do *backdoor* instalado, o atacante estabeleceu acesso remoto ao sistema, permitindo monitoramento contínuo e intervenções subsequentes sem necessidade de presença física.

É importante ressaltar que, embora esse cenário descreva ações maliciosas, o objetivo da análise é educacional e voltado para a conscientização sobre vulnerabilidades em sistemas ciberfísicos. A implementação de medidas de segurança adequadas, como autenticação, criptografia de comunicações, controle de acesso físico e monitoramento de atividades anômalas, pode mitigar os riscos descritos.

4.3.2 Implementação de Ataque *Man-in-the-Middle*

Com acesso privilegiado ao computador hospedeiro, o atacante procedeu à implementação de um ataque *man-in-the-middle* direcionado à comunicação serial entre MATLAB® e o microcontrolador ESP32. O objetivo dessa etapa é interceptar, registrar e, posteriormente, manipular os dados trocados entre o algoritmo de controle e a planta física, sem que o operador do sistema perceba qualquer anomalia.

4.3.2.1 Intercepção da Comunicação Serial

A comunicação serial entre MATLAB® e ESP32 ocorre utilizando uma porta COM virtual (por exemplo, COM3), criada pelo driver *USB-to-Serial* do sistema operacional *Windows*. Para interceptar essa comunicação, o atacante implementou um *software* intermediário que atua como *proxy* transparente entre o MATLAB® e o ESP32.

O procedimento de implementação envolveu as seguintes etapas:

1. **Identificação da porta serial original:** Utilizando ferramentas de linha de comando do *Windows* (como `mode` ou `wmic`), o atacante identificou que o ESP32 está conectado à porta COM3, operando a 115200 baud com configuração 8N1 (8 bits de dados, sem paridade, 1 bit de parada);
2. **Criação de par de portas seriais virtuais:** Com privilégios administrativos, o atacante modificou registros do sistema operacional para criar um par de portas seriais virtuais interconectadas (por exemplo, COM10 e COM11). Estas portas virtuais funcionam como um canal de comunicação bidirecional interno ao sistema operacional, onde dados escritos em uma porta aparecem automaticamente na outra;

3. **Redirecionamento da aplicação MATLAB®:** O atacante reconfigurou o sistema para que o MATLAB® se comunicasse com a porta virtual COM10 ao invés da porta física COM3. Essa modificação pode ser realizada utilizando diversas técnicas:

- (a) *Modificação direta do código-fonte:* Alteração da *string* de configuração da porta serial no script de controle (e.g., substituição de "COM3" por "COM10" na chamada da função `serialport`);
- (b) *Alteração de arquivos de configuração:* Modificação de parâmetros em arquivos externos (`.ini`, `.json`, `.mat`) que definem a porta de comunicação;
- (c) *Hijacking de função via manipulação do path:* Criação de uma função intermediária `serialport.m` em diretório com prioridade superior no caminho de busca do MATLAB®, que intercepta chamadas à porta COM3 e as redireciona transparentemente para COM10;
- (d) *Injeção de código via backdoor:* Execução de comandos por meio do acesso remoto previamente estabelecido para modificar dinamicamente o ambiente de execução antes da inicialização do script de controle.

4. **Implementação do proxy *man-in-the-middle*:** O atacante desenvolveu e executou um *software* intermediário que:

- Abre simultaneamente conexões com a porta virtual COM11 (conectada ao MATLAB® via COM10) e com a porta física COM3 (conectada ao ESP32);
- Retransmite todos os dados recebidos da porta COM11 para a porta COM3, e vice-versa, mantendo a comunicação aparentemente normal;
- Registra todos os dados transmitidos em ambas as direções em arquivos de *log* para análise posterior.

A Figura 30 ilustra a topologia normal de comunicação do sistema de controle, onde o MATLAB® se comunica diretamente com o microcontrolador ESP32 por meio da porta serial COM3. Nessa configuração, os dados de medição $y[k]$ e as ações de controle $u[k]$ trafegam sem intermediários entre o algoritmo de controle e a planta física.

Após a implementação do ataque *man-in-the-middle*, a topologia de comunicação é alterada conforme ilustrado na Figura 31. O *software* malicioso intercepta a comunicação utilizando portas seriais virtuais (COM10/COM11), retransmitindo os dados pela porta física COM3. Essa configuração permite ao atacante total visibilidade sobre o funcionamento do sistema de controle, incluindo os valores medidos dos sensores, as ações de controle calculadas pelo algoritmo, e a temporização da comunicação. Além disso, o atacante pode, a qualquer momento, modificar os dados em trânsito para injetar perturbações controladas ou implementar ataques.

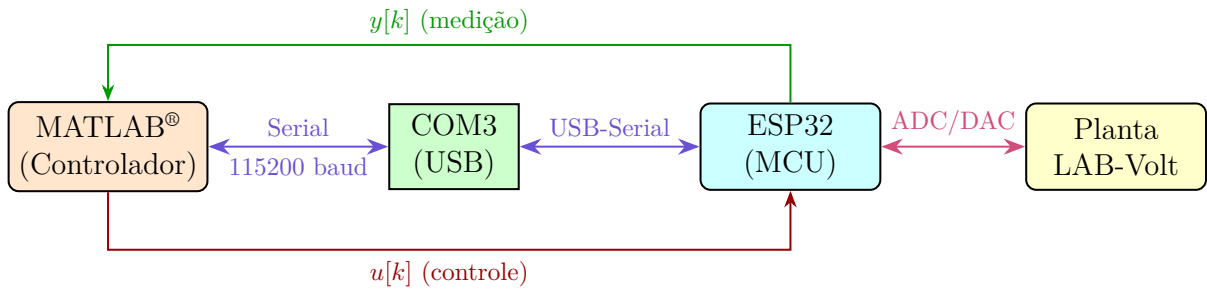


Figura 30 – Topologia normal de comunicação do sistema de controle, com conexão direta entre MATLAB® e ESP32 via porta serial COM3.

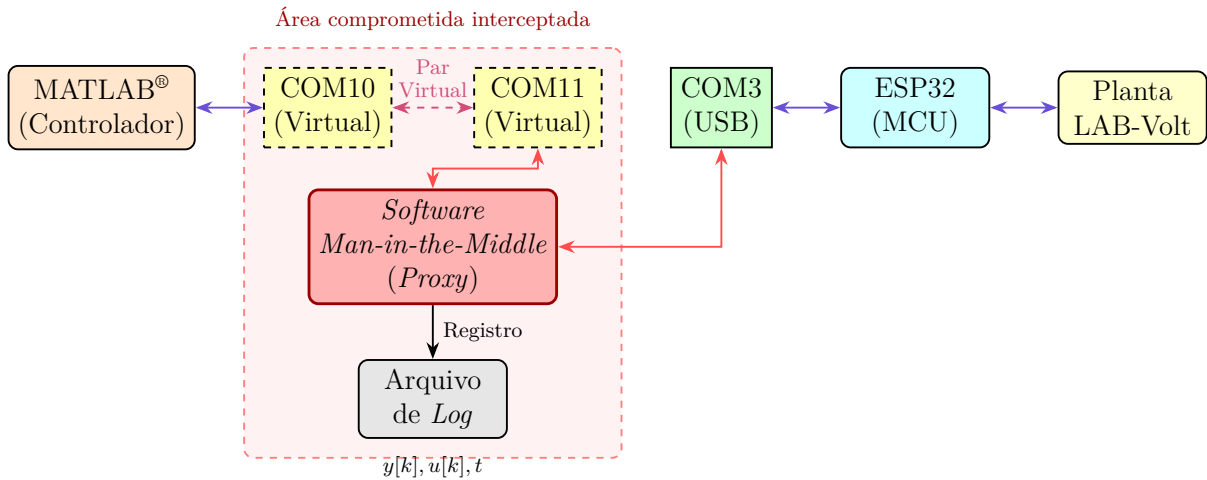


Figura 31 – Topologia de comunicação após implementação do ataque *man-in-the-middle*.

4.3.3 Identificação do Sistema Utilizando Excitação Persistente

Com o sistema *man-in-the-middle* operacional e capacidade de monitoramento passivo estabelecida, o atacante procedeu à fase seguinte: a identificação da dinâmica da válvula pneumática e sua interação com o sistema de controle de nível. O objetivo dessa etapa é obter um modelo matemático da planta que permita ao atacante projetar estratégias de ataque.

4.3.3.1 Injeção de Sinal de Excitação

Para realizar a identificação de sistema, é necessário que os dados coletados sejam persistentemente excitantes, ou seja, que contenham informação suficiente sobre a dinâmica do sistema em uma faixa adequada de frequências. Durante operação normal em regime permanente, com o nível estabilizado no *setpoint* desejado, a ação de controle varia minimamente, fornecendo dados pouco informativos para identificação.

Para contornar essa limitação, o atacante utilizou sua capacidade de manipulação dos dados para injetar uma perturbação de baixa amplitude na ação de controle. Especificamente, o *man-in-the-middle* foi modificado para adicionar um sinal de excitação $\delta u(t)$ à

ação de controle nominal $u_{nom}(t)$ calculada pelo MATLAB®:

$$u_{real}(t) = u_{nom}(t) + \delta u(t), \quad (4.41)$$

em que $u_{real}(t)$ é o sinal efetivamente transmitido ao ESP32 e aplicado à válvula pneumática.

O sinal de excitação $\delta u(t)$ foi projetado com as seguintes características:

- **Tipo:** Sequência binária pseudo-aleatória (PRBS, do inglês *pseudo-random binary sequence*), garantindo espectro de frequência amplo;
- **Amplitude:** Suficientemente pequena para não causar variações perceptíveis no nível do tanque que alertassem os operadores, mas suficientemente grande para excitar a dinâmica da válvula acima do nível de ruído de medição;
- **Duração:** Período de coleta de dados estendido (tipicamente várias constantes de tempo do sistema) para garantir a identificação adequada dos parâmetros do modelo.

4.3.3.2 Coleta e Registro de Dados

Durante o período de excitação, o *software* registrou simultaneamente:

- A ação de controle modificada $u_{real}(t)$ efetivamente aplicada à válvula pneumática;
- As medições do sensor de posição $p(t)$ retornadas pelo ESP32 ao MATLAB®;

A Figura 32 apresenta os sinais de entrada (ação de controle com perturbação PRBS) e saída (resposta do sensor) coletados durante o experimento de identificação. Observa-se que a amplitude da perturbação foi mantida em níveis que não comprometem a operação normal do sistema. O sinal de entrada $u_{real}(t)$ oscila entre aproximadamente 6,2 V e 6,6 V, representando uma variação de 0,4 V sobre um fundo de escala de 8 V, o que corresponde a 5% do *span* disponível. A saída $y(t)$ responde com variações entre aproximadamente 2,60 V e 2,78 V, representando uma excursão de 0,18 V sobre o fundo de escala útil de 1,8 V (zona de operação entre 1,5 V e 3,3 V, excluída a zona morta), correspondendo a aproximadamente 10% do *span* operacional disponível.

Os dados foram registrados com a mesma taxa de amostragem utilizada pelo sistema de controle original, sincronizados usando os *timestamps* das mensagens seriais. Ao término do período de coleta, o atacante desabilitou o sinal de excitação, permitindo que o sistema retornasse à operação normal sem vestígios da atividade maliciosa.

Os dados coletados foram então transferidos para um ambiente seguro controlado pelo atacante, onde as etapas de identificação de sistema e projeto do ataque de superamostragem seriam realizadas *offline*, sem risco de detecção.

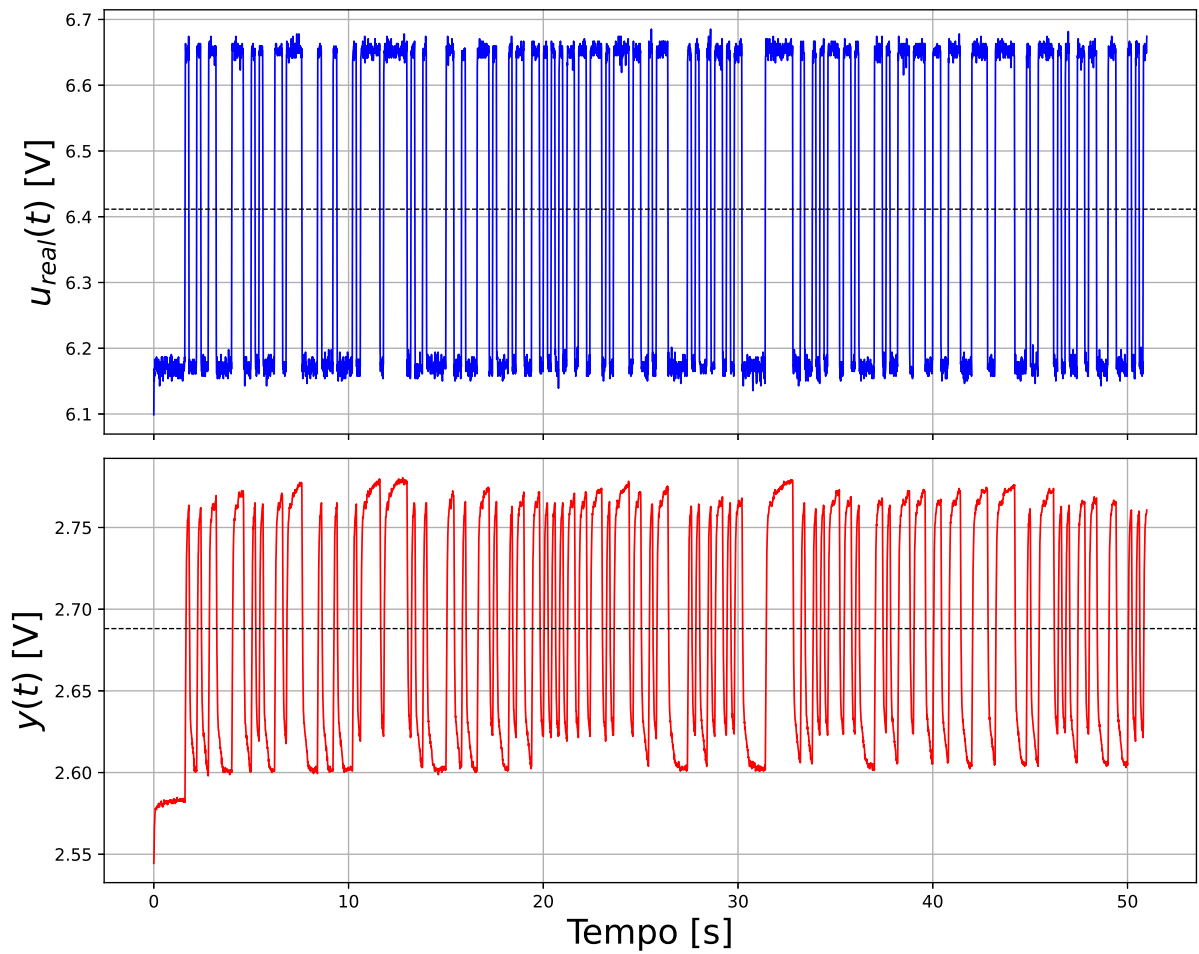


Figura 32 – Dados coletados durante o experimento de identificação: (a) sinal de entrada $u_{real}(t)$ com perturbação PRBS; (b) resposta do sensor de posição $y(t)$. A linha tracejada representa o ponto de operação do sistema, para o sinal de entrada é 6,412V e para a resposta do sensor de posição é 2,688V.

4.3.3.3 Metodologia de Identificação de Sistemas

Para a estimação do modelo matemático da planta, o atacante implementou um algoritmo de identificação de sistemas lineares, testando múltiplas estruturas de modelos candidatos. Os dados coletados foram divididos em conjunto de estimação (70%) e conjunto de validação (30%), seguindo boas práticas de modelagem preditiva (LJUNG, 1999).

Quatro estruturas de modelos lineares foram avaliadas:

- **ARX** (*AutoRegressive with eXogenous input*): Modelo autorregressivo com entrada exógena;
- **ARMAX** (*AutoRegressive Moving-Average with eXogenous input*): Modelo autorregressivo com média móvel e entrada exógena;
- **OE** (*Output-Error*): Modelo de erro de saída;
- **BJ** (*Box-Jenkins*): Modelo de *Box-Jenkins*.

Uma descrição detalhada de cada estrutura de modelo, pode ser encontrada em Aguirre (2015).

4.3.3.4 Validação do Modelo

A validação dos modelos candidatos foi realizada mediante critérios quantitativos e qualitativos. O índice de ajuste percentual (*fit*) quantifica a capacidade do modelo de reproduzir dados de validação não utilizados na estimação (LJUNG, 1999):

$$\text{Fit} = 100 \times \left(1 - \frac{\|y - \hat{y}\|}{\|y - \bar{y}\|} \right) \%, \quad (4.42)$$

em que y é a saída medida, $\hat{y}$ é a saída predita pelo modelo e $\bar{y}$ é a média da saída medida. A análise de resíduos verifica se o erro de predição $\varepsilon(k) = y(k) - \hat{y}(k)$ possui características de ruído branco, avaliando-se a função de autocorrelação dos resíduos. Para um modelo adequado, a autocorrelação deve estar dentro dos limites de confiança para todos os atrasos $\tau \neq 0$, indicando que os resíduos são não correlacionados e que o modelo capturou adequadamente a dinâmica do sistema.

4.3.3.5 Seleção do Modelo: Estrutura ARMAX

A Tabela 12 apresenta os resultados comparativos dos quatro modelos candidatos avaliados mediante o índice de ajuste percentual (*Fit*) calculado sobre os dados de validação.

Tabela 12 – Comparação dos modelos candidatos para identificação do sistema.

Modelo	<i>Fit</i> (%)
ARX	90,26
ARMAX	93,38
OE	75,71
BJ	90,66

O modelo ARMAX obteve o maior índice de ajuste, com *Fit* de 93,38%, superando todos os modelos candidatos. A validação visual do desempenho dos modelos é apresentada na Figura 33. Observa-se que o modelo ARMAX (linha vermelha tracejada) apresenta sobreposição quase precisa aos dados medidos (linha preta contínua), capturando adequadamente tanto os regimes transitórios quanto os de regime permanente. O modelo OE (linha verde tracejada), por sua vez, apresenta desvios mais notáveis, consistentes com seu menor índice de *Fit* (75,71%).

Para verificar a qualidade estatística dos modelos, a Figura 34 exhibe a análise de resíduos mediante a função de autocorrelação, que determina se o modelo capturou toda a dinâmica determinística do sistema. Idealmente, a autocorrelação deve permanecer dentro do intervalo de confiança de 95% (linhas pretas tracejadas). Observa-se que o modelo

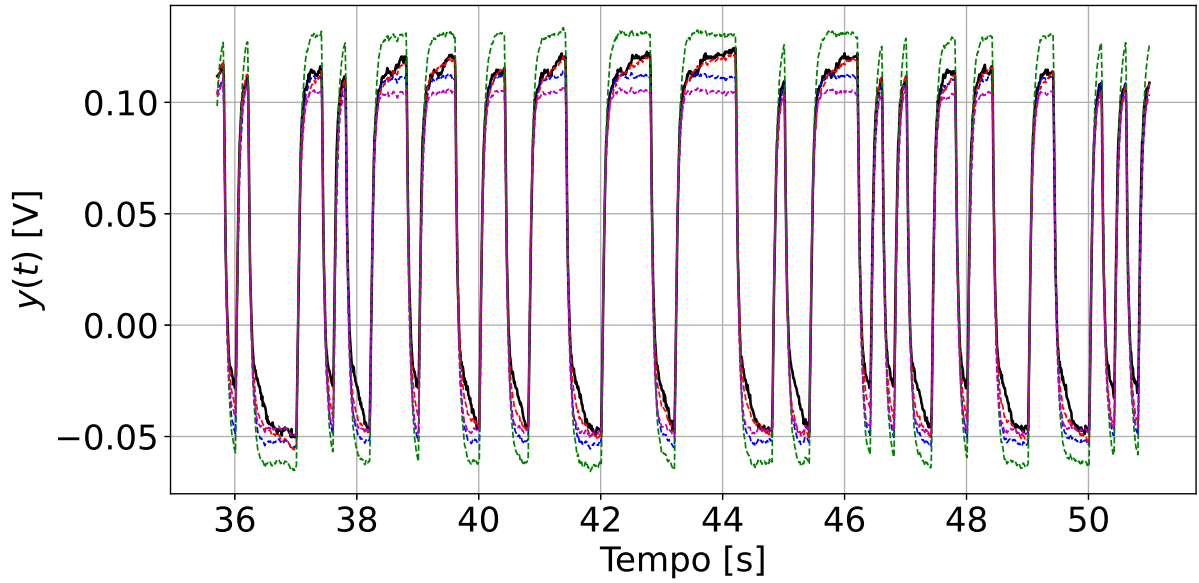


Figura 33 – Comparação entre a saída medida e as saídas simuladas pelos modelos identificados utilizando dados de validação. As curvas representam: Medido (preto, linha contínua), ARX (azul, tracejado), ARMAX (vermelho, tracejado), OE (verde, tracejado) e BJ (roxo, tracejado).

ARMAX (barras vermelhas) e BJ (barras roxas) apresentam comportamento próximo do ruído branco, com a maioria dos *lags* dentro dos limites de confiança.

Considerando o compromisso entre desempenho e complexidade computacional, o modelo ARMAX foi selecionado para a implementação do ataque. Sua estrutura requer apenas três polinômios (A , B e C) e entregou o maior índice de ajuste (93,38%), o que é crítico para minimizar o erro de estimação do sinal de controle durante o ataque.

A Figura 35 detalha o desempenho final do modelo ARMAX escolhido. O gráfico superior mostra a aderência do modelo aos dados de validação, enquanto o gráfico inferior mostra que o erro de modelagem permanece limitado e com magnitude reduzida em comparação à amplitude do sinal.

O modelo ARMAX de ordem (n_a, n_b, n_c) com atraso n_k é definido pela equação de diferenças (LJUNG, 1999):

$$A(q^{-1})y(k) = B(q^{-1})u(k - n_k) + C(q^{-1})e(k), \quad (4.43)$$

em que $y(k)$ é a saída do sistema no instante k , $u(k)$ é a entrada do sistema no instante k , $e(k)$ é o ruído branco com variância σ^2 , q^{-1} é o operador de atraso unitário tal que $q^{-1}y(k) = y(k - 1)$, e n_k é o atraso puro (*dead time*) do sistema.

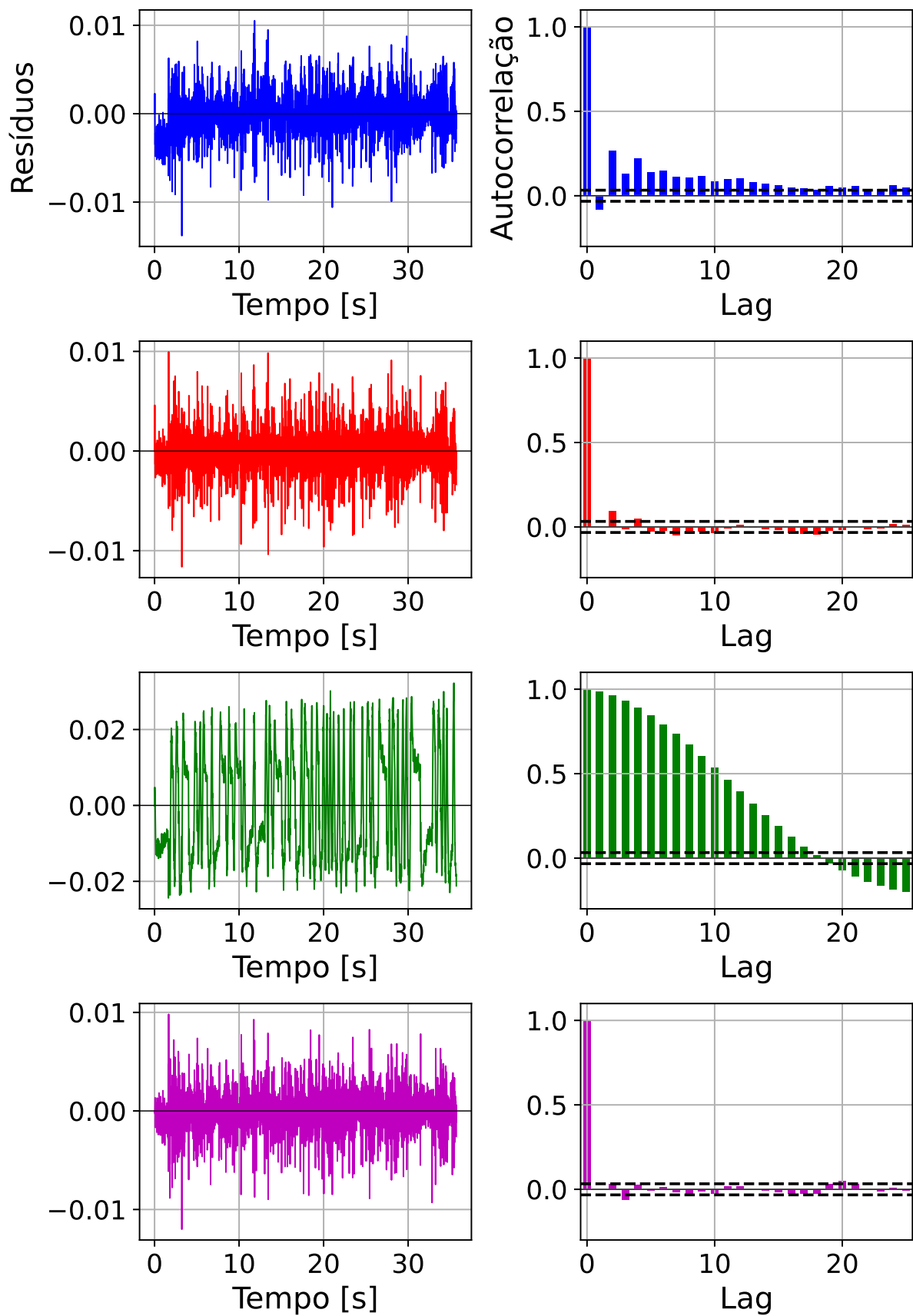


Figura 34 – Análise de resíduos (painel da esquerda) e autocorrelação dos resíduos (painel da direita). As cores correspondem aos modelos: ARX (azul), ARMAX (vermelho), OE (verde) e BJ (roxo). Nos gráficos de autocorrelação, as linhas horizontais pretas tracejadas delimitam o intervalo de confiança de 95%.

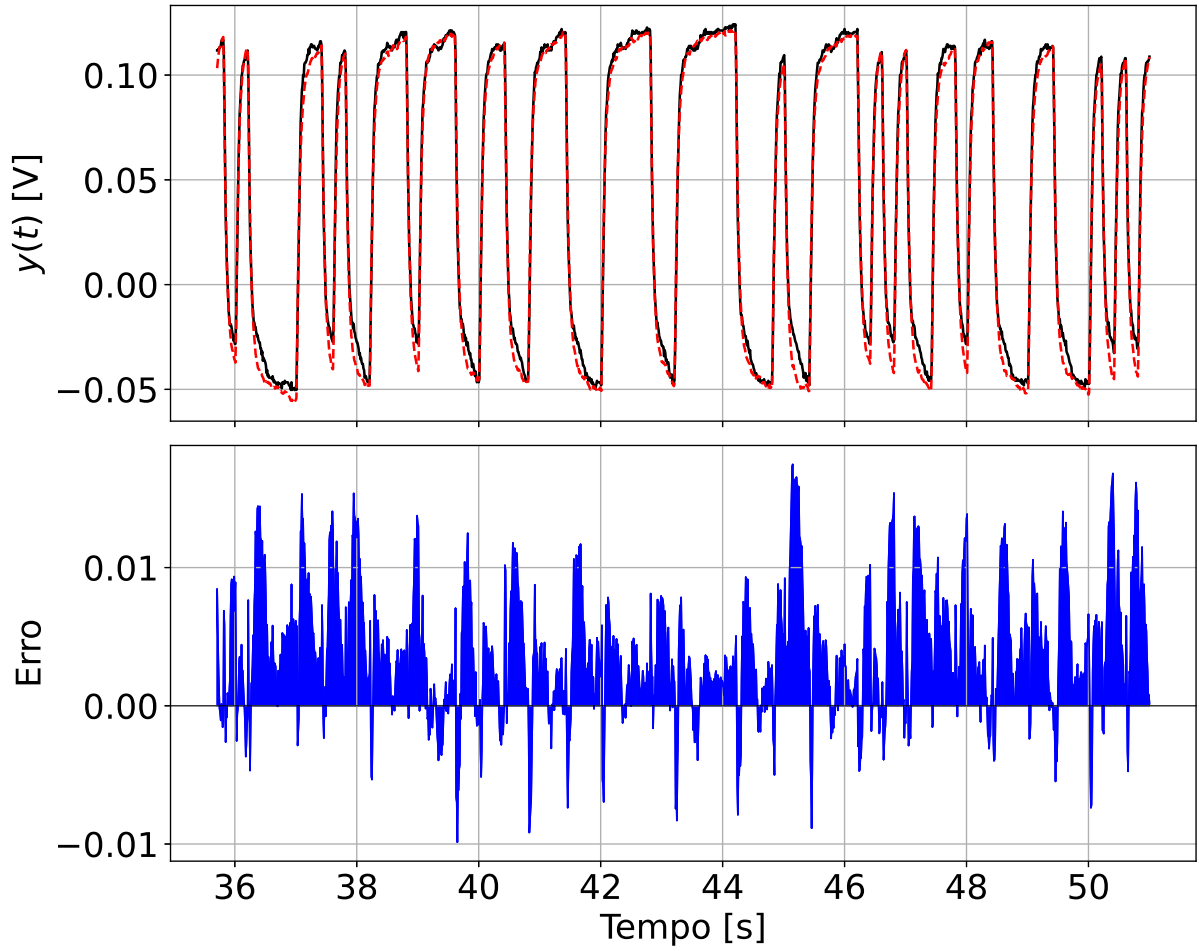


Figura 35 – Validação do modelo ARMAX escolhido. O gráfico superior compara os dados de validação (preto, linha contínua) com a resposta do modelo ARMAX (vermelho, tracejado). O gráfico inferior apresenta a evolução do erro de modelagem ao longo do tempo.

Os polinômios $A(q^{-1})$, $B(q^{-1})$ e $C(q^{-1})$ são definidos como:

$$A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} + \dots + a_{n_a} q^{-n_a}, \quad (4.44)$$

$$B(q^{-1}) = b_1 q^{-1} + b_2 q^{-2} + \dots + b_{n_b} q^{-n_b}, \quad (4.45)$$

$$C(q^{-1}) = 1 + c_1 q^{-1} + c_2 q^{-2} + \dots + c_{n_c} q^{-n_c}. \quad (4.46)$$

O polinômio $A(q^{-1})$ representa a dinâmica autorregressiva do sistema, $B(q^{-1})$ modela a influência da entrada na saída, e $C(q^{-1})$ caracteriza a estrutura do ruído conforme um modelo de média móvel. Essa última característica é importante, pois permite que o modelo ARMAX capture perturbações correlacionadas temporalmente, resultando em estimativas de parâmetros mais precisas quando o ruído do processo não é puramente branco.

A função de transferência do modelo ARMAX, relacionando entrada e saída na

ausência de ruído, é dada por:

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})} q^{-n_k}. \quad (4.47)$$

Para o sistema identificado, foram utilizadas ordens $n_a = 2$, $n_b = 2$, $n_c = 2$ e atraso $n_k = 1$, resultando em um modelo de segunda ordem com dinâmica de ruído modelada. Essa escolha se baseou no comportamento do sistema, que se assemelha a um de segunda ordem, e na constatação de que ordens superiores não proporcionaram aumento significativo no índice de ajuste, justificando a opção por simplicidade.

4.3.3.6 Parâmetros do Modelo ARMAX Identificado

A partir do procedimento de estimação, os coeficientes dos polinômios do modelo ARMAX foram determinados. Para o sistema identificado, os polinômios resultantes são:

$$A(q^{-1}) = 1 - 1,7627 q^{-1} + 0,7683 q^{-2}, \quad (4.48)$$

$$B(q^{-1}) = 0,0659 q^{-1} - 0,0638 q^{-2}, \quad (4.49)$$

$$C(q^{-1}) = 1 - 0,8823 q^{-1} - 0,0003 q^{-2}, \quad (4.50)$$

com variância do ruído branco estimada em $\sigma_e^2 = 2,07 \times 10^{-4}$ e período de amostragem $T_s = 0,01$ s.

Substituindo os coeficientes na Equação (4.43), obtém-se a representação entrada-saída do sistema identificado:

$$\begin{aligned} y(k) = & 1,7627 y(k-1) - 0,7683 y(k-2) \\ & + 0,0659 u(k-1) - 0,0638 u(k-2) \\ & + e(k) - 0,8823 e(k-1) - 0,0003 e(k-2). \end{aligned} \quad (4.51)$$

A função de transferência do modelo, desconsiderando o modelo do ruído, é dada por:

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})} = \frac{0,0659 q^{-1} - 0,0638 q^{-2}}{1 - 1,7627 q^{-1} + 0,7683 q^{-2}}. \quad (4.52)$$

Alternativamente, em termos do operador de avanço q , a função de transferência pode ser expressa como:

$$G(q) = \frac{0,0659 q - 0,0638}{q^2 - 1,7627 q + 0,7683}. \quad (4.53)$$

4.3.3.7 Conversão para Espaço de Estados

Para o projeto do filtro de *Kalman* superamostrado e implementação do ataque, é necessário converter o modelo ARMAX para a representação em espaço de estados. A

conversão é realizada utilizando da forma canônica observável, adequada para sistemas onde a saída é diretamente medida.

Considerando a Função de Transferência (4.53) com polinômio denominador $A(q) = q^2 + a_1q + a_2$ e numerador $B(q) = b_1q + b_2$, em que $a_1 = -1,7627$, $a_2 = 0,7683$, $b_1 = 0,0659$ e $b_2 = -0,0638$, a representação em espaço de estados na forma canônica observável é obtida definindo-se o vetor de estados $\mathbf{x}(k) = [x_1(k) \ x_2(k)]^\top$ tal que:

$$x_1(k+1) = -a_2 x_2(k) + b_2 u(k), \quad (4.54)$$

$$x_2(k+1) = x_1(k) - a_1 x_2(k) + b_1 u(k), \quad (4.55)$$

$$y(k) = x_2(k). \quad (4.56)$$

Organizando na forma matricial padrão:

$$\mathbf{x}(k+1) = \mathbf{A}_D \mathbf{x}(k) + \mathbf{b}_D u(k), \quad (4.57)$$

$$y(k) = \mathbf{c}_D \mathbf{x}(k) + d_D u(k), \quad (4.58)$$

em que as matrizes do sistema discreto são:

$$\mathbf{A}_D = \begin{bmatrix} 0 & -a_2 \\ 1 & -a_1 \end{bmatrix} = \begin{bmatrix} 0 & -0,7683 \\ 1 & 1,7627 \end{bmatrix}, \quad (4.59)$$

$$\mathbf{b}_D = \begin{bmatrix} b_2 \\ b_1 \end{bmatrix} = \begin{bmatrix} -0,0638 \\ 0,0659 \end{bmatrix}, \quad (4.60)$$

$$\mathbf{c}_D = [0 \ 1], \quad d_D = 0. \quad (4.61)$$

4.3.4 Implementação dos Ataques Cibernéticos

Com o acesso privilegiado obtido mediante *software man-in-the-middle* e com o sistema identificado, o atacante possui a capacidade de interceptar, modificar e injetar dados no canal de comunicação entre o controlador e a planta. Essa posição estratégica permite a implementação de diversos tipos de ataques cibernéticos.

As implementações práticas e análises experimentais dos ataques clássicos descritos nas Seções 2.1 (Ataques de Injeção de Dados Falsos), 2.2 (Ataques de Repetição) e 2.3 (Ataques de Atraso Temporal) encontram-se detalhadas no Apêndice B. Nesta seção, apresentam-se os resultados da implementação do ataque de superamostragem, detalhado no Capítulo 3, no sistema identificado.

4.3.4.1 Ataque Simulado (Rastreamento)

Para validar a metodologia proposta, o ataque de superamostragem foi implementado em simulação utilizando o modelo discreto identificado do sistema. A Tabela 13 apresenta os parâmetros utilizados na simulação.

Tabela 13 – Parâmetros da simulação do ataque de superamostragem.

Parâmetro	Valor
<i>Matrizes do sistema discreto</i>	
$\mathbf{A}_D$	$\begin{bmatrix} 0 & -0,7683 \\ 1 & 1,7627 \end{bmatrix}$
$\mathbf{b}_D$	$\begin{bmatrix} -0,0638 \\ 0,0659 \end{bmatrix}$
$\mathbf{b}_w$	$\begin{bmatrix} -0,7687 \\ 0,8805 \end{bmatrix}$
$\mathbf{c}_D$	$\begin{bmatrix} 0 & 1 \end{bmatrix}$
<i>Parâmetros de ruído</i>	
Variância do ruído (σ_e^2)	$2,07 \times 10^{-4} \text{ V}^2$
Desvio padrão (σ_e)	$0,0144 \text{ V}$
<i>Parâmetros de simulação</i>	
Condição inicial ($\mathbf{x}_0$)	$\begin{bmatrix} 0 & 0 \end{bmatrix}^\top$
Tempo de simulação (t_{sim})	$1,5 \text{ s}$
Período de amostragem (T_D)	10 ms
Fator de superamostragem (r)	4
Período superamostrado (T_d)	$2,5 \text{ ms}$
Setpoint (r)	$3,0 \text{ V}$
<i>Limites de saturação</i>	
u_{min}	0 V
u_{max}	15 V
<i>Ponderações do LQR</i>	
$\mathbf{Q}$	$\mathbf{I}_2$
R	1
Q_{int}	10

O controlador DLQR com ação integral foi projetado utilizando o método do sistema aumentado, resultando no ganho:

$$\mathbf{k}_{track} = [\mathbf{k}_x \ k_i] = [14,80 \ 21,58 \ -2,46], \quad (4.62)$$

em que $\mathbf{k}_x$ corresponde ao ganho de realimentação de estados e k_i ao ganho integral.

A Figura 36 apresenta a evolução temporal da saída do sistema para os cenários com e sem ataque.

A Figura 37 mostra a ação de controle nos dois cenários. No caso sem ataque, o sinal de controle é constante por partes, mudando apenas nos instantes de amostragem. No cenário com ataque, a ação de controle varia a cada $T_d = 2,5 \text{ ms}$, com valores otimizados para maximizar o desvio entre amostras enquanto satisfazem a condição de furtividade.

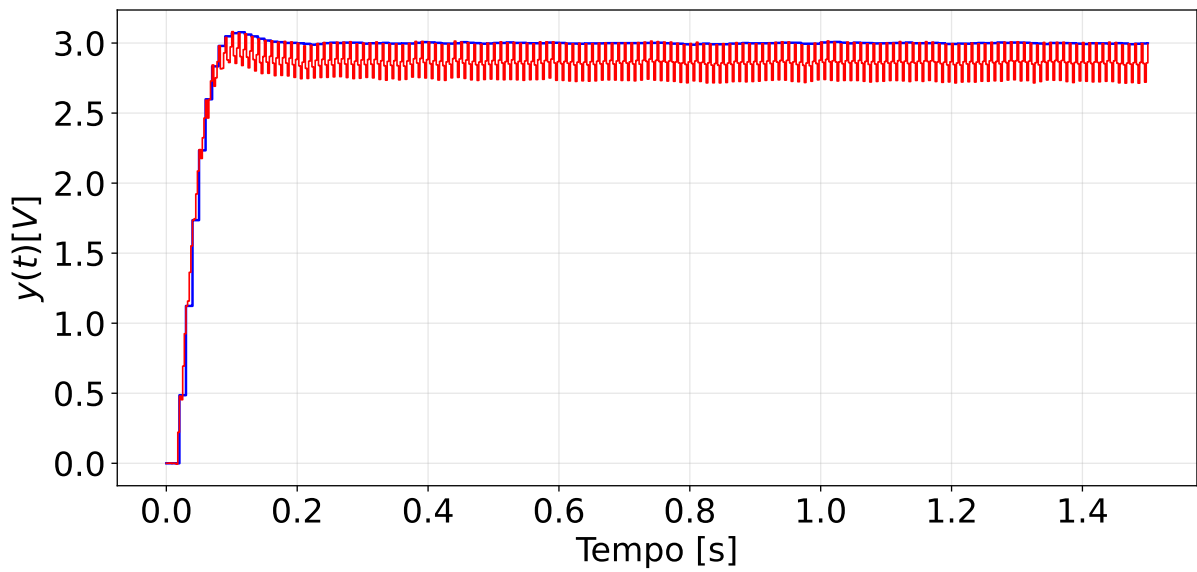


Figura 36 – Evolução temporal da saída para o problema de rastreamento com presença de ruído. Em azul sólido, o sistema sem ataque com $T_D = 10$ ms; em vermelho sólido, o sistema sob ataque de superamostragem com $T_d = 2,5$ ms.

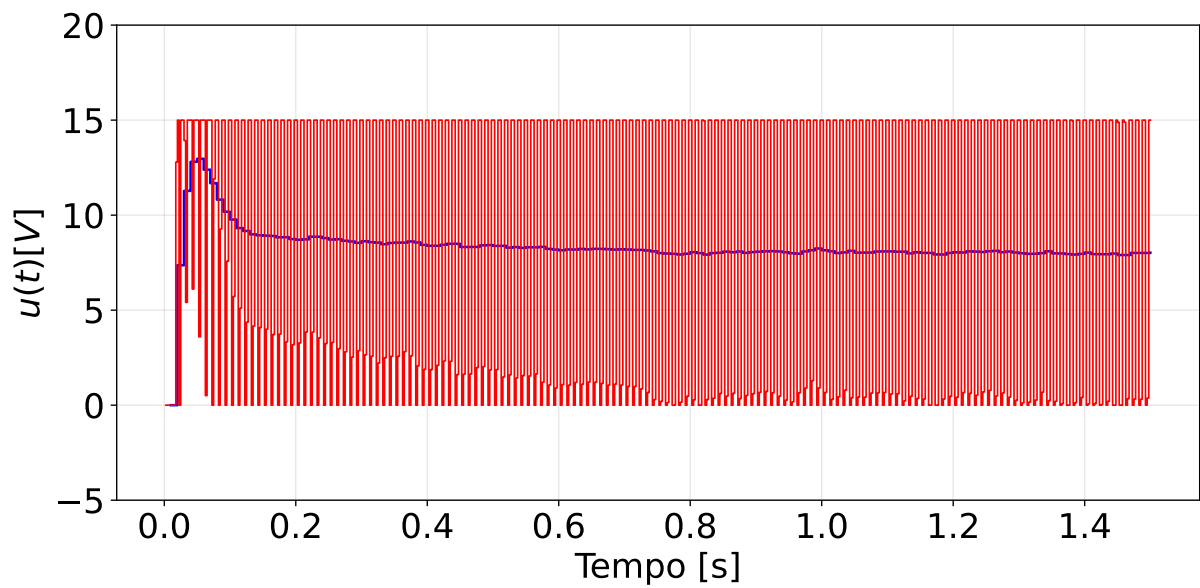


Figura 37 – Ação de controle para o problema de rastreamento com presença de ruído. Para mais informações, ver a legenda da Figura 36.

4.3.4.2 Análise de Impacto

A Tabela 14 apresenta as métricas de impacto do ataque de superamostragem. A energia RMS da ação de controle aumentou de 8,42 V para 10,71 V, representando amplificação de $1,27\times$ e incremento percentual de 27,26%. O impacto mais significativo é observado no índice de desgaste do atuador: o ataque elevou o valor de 22,79 V para 4423,75 V, resultando em razão de desgaste de $194,06\times$.

Tabela 14 – Métricas de impacto do ataque de superamostragem.

Métrica	Sem ataque	Com ataque
<i>Energia RMS – Equação (3.14)</i>		
E_{RMS} [V]	8,42	10,71
Razão de amplificação		$1,27\times$
<i>Taxa de Incremento Energético – Equação (3.15)</i>		
ΔE (%)		27,26%
<i>Índice de Desgaste do Atuador – Equação (3.16)</i>		
TV [V]	22,79	4423,75
Razão de desgaste		$194,06\times$

4.3.4.3 Ataque de Superamostragem Experimental

A validação experimental do ataque de superamostragem foi realizada na planta piloto LAB-Volt, utilizando a infraestrutura *man-in-the-middle* descrita na Subseção 4.3.2. O *software* de ataque implementa o algoritmo de otimização em tempo real, calculando a sequência de controle superamostrada $\{u_0, u_1, u_2, u_3\}$ a cada período de amostragem do controlador.

Adaptações para Implementação em Planta Física

Devido às limitações práticas da planta física, os limites de tensão de comando foram definidos em [3,0 V; 5,0 V], mais restritivos que os limites teóricos [0 V; 15 V], garantindo operação segura da válvula pneumática e evitando danos aos componentes hidráulicos.

Configuração Experimental

A Tabela 15 apresenta os parâmetros utilizados nos experimentos de validação do ataque de superamostragem na planta física.

O controlador nominal foi projetado via DLQR sobre o modelo linearizado da planta em torno da condição de operação. As matrizes de ponderação foram sintonizadas

Tabela 15 – Parâmetros dos experimentos de validação física do ataque de superamostragem.

Parâmetro	Valor
<i>Parâmetros temporais</i>	
Período de amostragem (T_D)	10 ms
Período superamostrado (T_d)	2,5 ms
Fator de superamostragem (r)	4
<i>Condições de operação</i>	
Condição inicial (y_0)	2,00 V
Setpoint (y_{ref})	3,00 V
<i>Limites de saturação</i>	
u_{min}	3,0 V
u_{max}	5,0 V
<i>Ponderações do controlador LQR</i>	
$\mathbf{Q}$	$\mathbf{I}_2$
R	10
Q_{int}	1

para obter resposta rápida sem sobressinal, priorizando a regulação dos estados físicos em relação ao esforço de controle.

Protocolo de Execução

Para garantir comparabilidade entre os cenários normal e atacado, foi estabelecido o seguinte protocolo experimental:

1. **Sincronização das condições iniciais:** Antes de cada experimento, a planta foi estabilizada na condição inicial $y_0 \approx 2,0$ V mediante a aplicação do sinal de controle mínimo ($u = 3,0$ V) por tempo superior a 10 s;
2. **Aquisição sequencial:** O experimento com controle normal foi executado primeiro, seguido pelo experimento sob ataque, ambos partindo de condições iniciais equivalentes;
3. **Implementação do ataque:** Durante o experimento atacado, o *software* malicioso interceptou cada comando $u^*[k]$ emitido pelo controlador LQR, calculou a sequência ótima $\mathbf{u}_a = [u_0, u_1, u_2, u_3]^\top$ por meio do algoritmo de otimização apresentado na Seção 3.2, e aplicou esta sequência à planta.

4.3.4.4 Resultados Experimentais

A Figura 38 apresenta a evolução temporal da saída da válvula para os cenários normal e atacado.

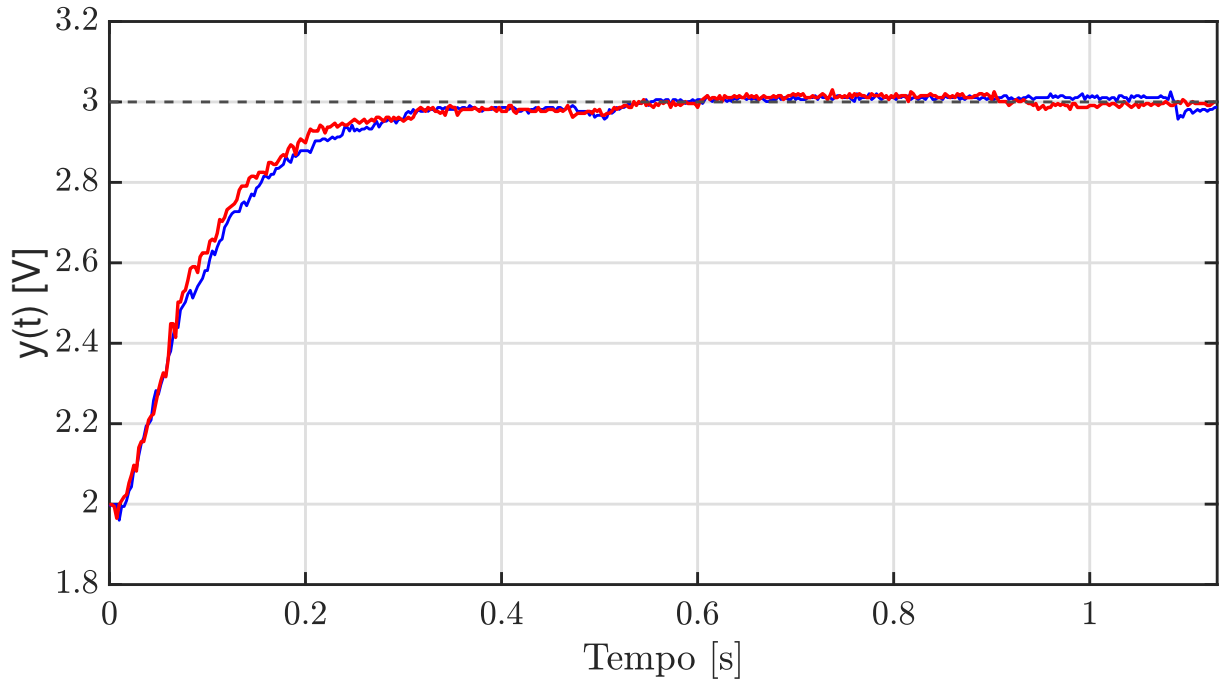


Figura 38 – Evolução temporal da tensão de saída da válvula nos experimentos físicos. Em azul, o sistema sob controle normal. Em vermelho, o sistema sob ataque de superamostragem. A linha tracejada preta indica o *setpoint* de 3,0 V.

Ambos os sistemas partem de aproximadamente 2,0 V e convergem para o *setpoint* de 3,0 V. A resposta transitória apresenta comportamento similar nos dois casos, com tempo de acomodação de aproximadamente 0,5 s. Observa-se que a saída sob ataque (vermelho) apresenta oscilações de pequena amplitude em torno da trajetória nominal (azul), porém as duas curvas permanecem próximas ao longo de todo o experimento. Em regime permanente, ambas estabilizam no *setpoint* com pequenas flutuações devido ao ruído de medição.

A Figura 39 ilustra as ações de controle aplicadas à planta nos dois experimentos.

O sinal de controle nominal (azul) apresenta variação suave, partindo de valores próximos ao limite inferior durante o transitório inicial e convergindo para aproximadamente 4,5 V em regime permanente.

O sinal de controle sob ataque (vermelho) apresenta oscilações de grande amplitude ao longo de todo o experimento. Durante o transitório inicial, o sinal alterna rapidamente entre os extremos. Em regime permanente, as oscilações persistem, com o sinal atingindo frequentemente o limite superior de saturação.

Análise Quantitativa de Impacto

A Tabela 16 resume as métricas de impacto do ataque calculadas a partir dos dados experimentais.

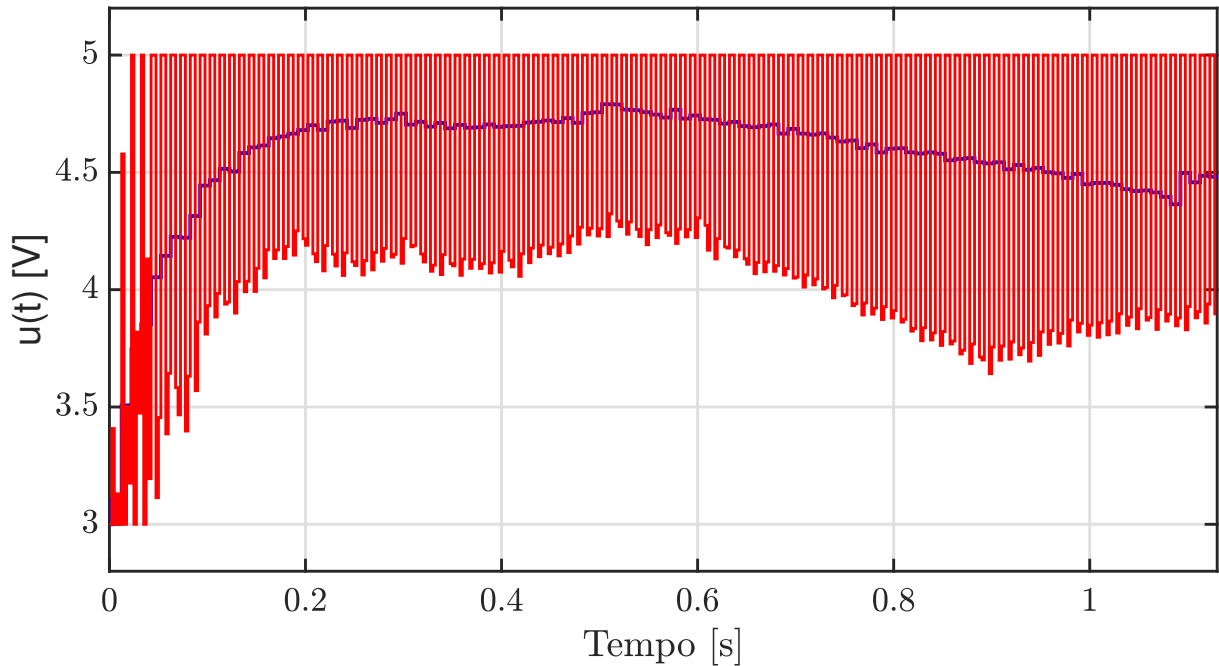


Figura 39 – Ação de controle (tensão de comando da válvula) aplicada nos experimentos físicos. Em azul, o controle nominal. Em vermelho, o controle de ataque.

Tabela 16 – Métricas de impacto do ataque de superamostragem nos experimentos físicos.

Métrica	Sem ataque	Com ataque
<i>Energia RMS – Equação (3.14)</i>		
E_{RMS} [V]	4,56	4,50
Razão de amplificação		0,99×
<i>Taxa de Incremento Energético – Equação (3.15)</i>		
ΔE (%)		−1,43%
<i>Índice de Desgaste do Atuador – Equação (3.16)</i>		
TV [V]	3,27	241,80
Razão de desgaste		73,94×

A energia RMS do sinal de controle apresentou variação mínima (−1,43%, razão de 0,99×), indicando que o ataque não aumentou a energia total aplicada, mas redistribuiu essa energia no tempo. Por outro lado, o índice de desgaste do atuador apresentou aumento: o controle atacado registrou TV de 241,80 V, representando razão de desgaste de 73,94× e incremento de 7295,03% em relação ao controle normal (3,27 V). Este resultado mostra o principal impacto físico do ataque: a indução de desgaste acelerado nos componentes mecânicos devido ao chaveamento de alta frequência entre os instantes de amostragem. Enquanto a energia total permanece praticamente inalterada, o padrão temporal das ações de controle é modificado, forçando o atuador a realizar transições bruscas e repetitivas que aceleram substancialmente a fadiga mecânica do sistema.

4.3.4.5 Detecção do Ataque

Para validar a capacidade de detecção do ataque em condições experimentais, foi implementado um Filtro de *Kalman* com o dobro da taxa de amostragem do sistema nominal. O ataque foi iniciado em $t = 0,5$ s, permitindo avaliar tanto a taxa de falsos positivos durante operação normal quanto a capacidade de detecção efetiva após a ativação do ataque.

O método de detecção baseado em autocorrelação das inovações do filtro utiliza uma janela deslizante de $W = 40$ amostras. O limiar foi calculado conforme a Equação (3.33), utilizando nível de confiança $\alpha = 0,001$, resultando em $\Psi_{ACF} \approx 0,520$.

A Figura 40 apresenta os resultados da detecção aplicada aos dados experimentais.

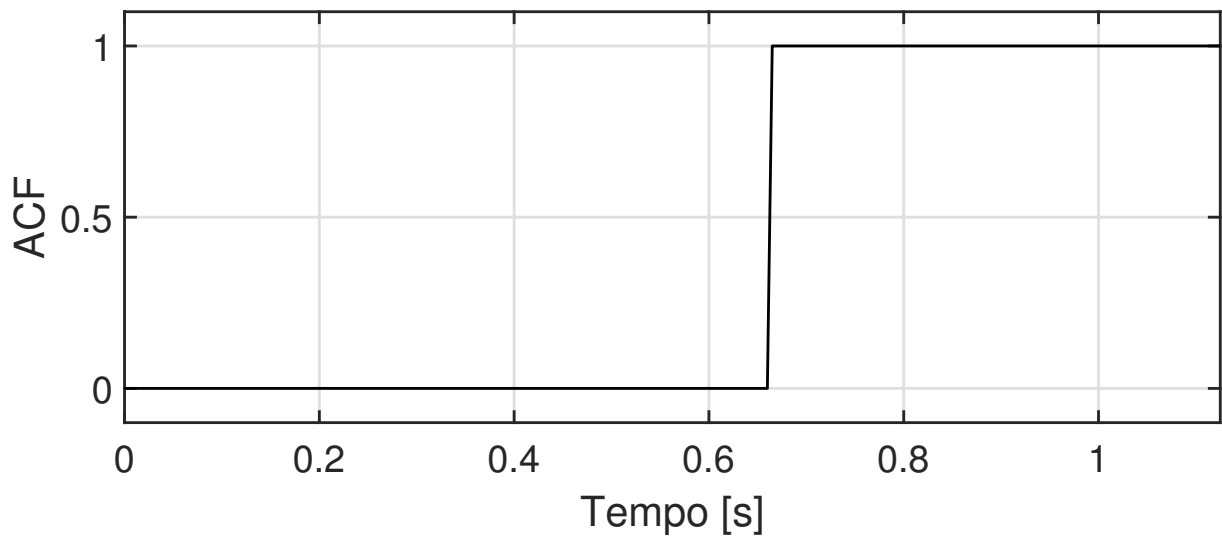


Figura 40 – Resultados da detecção do ataque de superamostragem nos experimentos físicos utilizando a função de autocorrelação.

Observa-se que o detector permanece inativo (valor 0) durante o intervalo de operação normal ($t < 0,5$ s), confirmando baixa taxa de falsos positivos. Após a ativação do ataque em $t = 0,5$ s, o índice de detecção transita para o estado ativo (valor 1) aproximadamente em $t \approx 0,63$ s, sinalizando a presença de correlações temporais anômalas nas inovações do filtro. O atraso de aproximadamente 0,1 s entre o início do ataque e sua detecção é inerente ao método baseado em janela deslizante, a janela de $W = 40$ amostras (correspondente a 0,2 s na taxa do detector) requer que dados suficientes sob ataque sejam acumulados para que a estatística de autocorrelação ultrapasse o limiar estabelecido.

Para fins de robustez, foi também avaliado o desempenho do detector baseado na estatística qui-quadrado das inovações, cujos resultados são apresentados na Figura 41.

Conforme ilustrado na Figura 41, o índice $g(j)$ apresenta picos elevados ainda durante a fase de operação normal (antes de $t = 0,5$ s), caracterizando falsos positivos que inviabilizam a utilização direta do método nesse cenário experimental. Esse comporta-

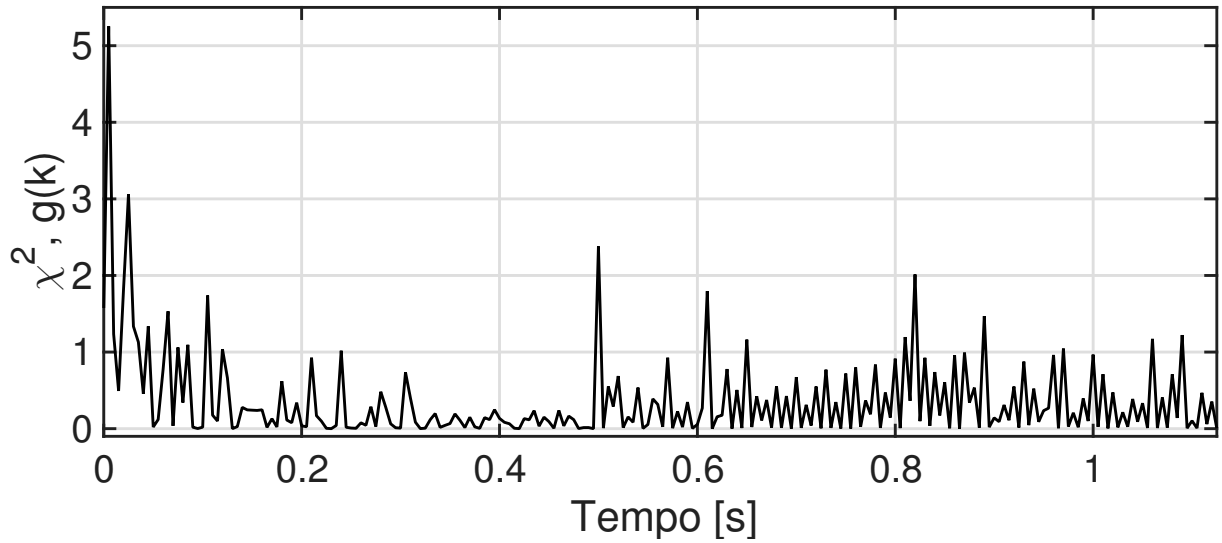


Figura 41 – Resultados da detecção do ataque de superamostragem nos experimentos físicos utilizando a estatística qui-quadrado χ^2 das inovações do filtro de *Kalman*.

mento decorre das perturbações e ruídos presentes na planta física, cujas características estatísticas diferem das assumidas pelo modelo do filtro, fazendo com que as inovações violem esporadicamente qualquer limiar razoável mesmo sem a presença do ataque. Adicionalmente, a definição do limiar Ψ_{χ^2} é intrinsecamente difícil nesse contexto: ao contrário do método baseado na ACF, cuja banda de confiança possui fundamentação estatística direta, o limiar qui-quadrado depende das propriedades do ruído real da planta, que raramente segue com precisão a distribuição gaussiana assumida. Por esses motivos, o método qui-quadrado não foi adotado como detector principal neste cenário, sendo o método baseado na ACF o escolhido para a detecção do ataque.

5 Conclusões

Essa dissertação investigou vulnerabilidades temporais em sistemas de controle digital, com foco no *ataque de superamostragem* e em estratégias para sua detecção. O trabalho demonstrou que a segurança de sistemas ciberfísicos não pode ser avaliada apenas nos instantes de amostragem do controlador, sendo necessário considerar o comportamento do sistema nos intervalos entre amostras, onde ações maliciosas podem ocorrer sem deixar rastros nos mecanismos convencionais de monitoramento.

5.1 Síntese do Trabalho

O Capítulo 1 contextualizou o cenário de vulnerabilidades em sistemas ciberfísicos, apresentando casos históricos como o *Maroochy Water Breach*, o *Stuxnet* e os ataques à infraestrutura elétrica da Ucrânia, que evidenciaram a necessidade de abordagens de segurança que considerem a natureza híbrida desses sistemas. Foram discutidos os principais tipos de ataque (FDI, *replay*, *delay*, *GPS spoofing* e espionagem) e as estratégias de defesa consolidadas na literatura.

O Capítulo 2 apresentou a fundamentação teórica, detalhando a modelagem matemática dos ataques de injeção de dados falsos, repetição, atraso temporal e espionagem. Para cada tipo de ataque, foram formalizadas as condições de furtividade e as estratégias de defesa correspondentes.

O Capítulo 3 desenvolveu a formulação matemática completa do ataque de superamostragem. Para sistemas lineares, o problema foi modelado como uma otimização sob restrições de furtividade (coincidência das saídas nos instantes de amostragem) e maximização do impacto energético. A extensão para sistemas MIMO foi apresentada, incluindo a análise de existência e unicidade de soluções. Para sistemas não lineares, duas abordagens complementares foram desenvolvidas: a modelagem *fuzzy Takagi-Sugeno*, que preserva a interpretabilidade física por meio da interpolação de subsistemas lineares locais, e a teoria dos operadores de *Koopman*, que eleva o sistema para um espaço de observáveis onde a dinâmica se torna globalmente linear. O capítulo também detalhou o método de detecção baseado em filtro de *Kalman* superamostrado, utilizando o detector estatístico χ^2 e a análise da função de autocorrelação como ferramentas de identificação.

O Capítulo 4 apresentou os resultados experimentais organizados em três cenários de complexidade crescente. O Cenário 1 validou computacionalmente o ataque em sistemas lineares, incluindo o integrador duplo e o sistema MIMO da aeronave F-8C DFBW, demonstrando incrementos energéticos superiores a 1900% enquanto o ataque permanecia

indetectável pelos métodos convencionais. O Cenário 2 estendeu a validação para sistemas não lineares utilizando um motor de corrente contínua com saturação magnética, em que as abordagens *Takagi-Sugeno* e *Koopman* foram comparadas. O Cenário 3 apresentou a implementação física na planta piloto *LAB-Volt 3503-M0*, adotando a perspectiva completa do atacante desde o reconhecimento do sistema até a execução do ataque, incluindo a implementação de um *man-in-the-middle* para interceptação da comunicação serial e a validação experimental de múltiplos tipos de ataques ciberfísicos.

5.2 Principais Contribuições

As contribuições dessa dissertação podem ser organizadas em três eixos principais:

Contribuições teóricas: O trabalho formalizou matematicamente o ataque de superamostragem como um problema de otimização com restrições de igualdade (condição de coincidência) e desigualdade (limites de saturação). A análise de existência e unicidade de soluções foi desenvolvida para sistemas lineares SISO e estendida para sistemas MIMO, estabelecendo condições necessárias e suficientes baseadas nas propriedades da matriz de controlabilidade do ataque. A extensão para sistemas não lineares via *Takagi-Sugeno* e operadores de *Koopman* representa uma contribuição original que amplia o escopo de aplicação do ataque para uma classe mais ampla de sistemas.

Contribuições metodológicas: O método de detecção baseado em observador superamostrado com detector χ^2 e análise de autocorrelação foi desenvolvido e validado. O algoritmo GRASP foi adaptado para a otimização do sinal de ataque, demonstrando desempenho superior às abordagens baseadas em programação semidefinida em termos de custo computacional e qualidade das soluções.

Contribuições experimentais: A validação em planta física real representa uma contribuição significativa, demonstrando a aplicabilidade prática das formulações desenvolvidas. Os experimentos na *LAB-Volt 3503-M0* confirmaram tanto a capacidade destrutiva do ataque (incremento de 7295% na variação total do sinal de controle) quanto a viabilidade de implementação em sistemas industriais reais. A demonstração completa do vetor de ataque, desde o reconhecimento até a execução, fornece um *framework* para avaliação de vulnerabilidades em sistemas de controle.

5.3 Limitações Identificadas

Apesar dos resultados obtidos, algumas limitações devem ser consideradas:

Requisitos de conhecimento do modelo: O ataque de superamostragem requer que o atacante possua um modelo suficientemente preciso da planta. Embora o Cenário 3 tenha demonstrado que esse modelo pode ser obtido utilizando técnicas de identificação de

sistemas, a qualidade do ataque depende diretamente da fidelidade do modelo identificado. Sistemas com dinâmicas altamente não lineares ou variantes no tempo podem apresentar maior robustez natural ao ataque.

Requisitos de aquisição de dados: A estratégia de detecção baseada no filtro de *Kalman* superamostrado requer que o sistema de monitoramento seja capaz de adquirir medições dos sensores em uma frequência superior à do controlador original. Em aplicações industriais onde o *hardware* de aquisição opera na mesma taxa do controlador, essa condição pode não ser satisfeita sem a adição de infraestrutura dedicada, o que representa um custo adicional de implementação.

Restrições de atuação: O ataque está restrito/confinado aos limites físicos de saturação do atuador. Em sistemas com faixas de atuação estreitas ou com atuadores de resposta lenta, o impacto do ataque pode ser significativamente reduzido. Os experimentos na planta *LAB-Volt* utilizaram limites seguros (3,0 V a 5,0 V) que limitaram o potencial destrutivo do ataque.

Sensibilidade dos métodos de detecção a condições experimentais: Os resultados experimentais na planta *LAB-Volt* mostraram que o desempenho dos métodos de detecção pode ser sensível às características do sistema físico. O detector baseado na estatística qui-quadrado, embora eficaz em simulações, apresentou desempenho limitado nos experimentos práticos, exibindo alta taxa de falsos positivos durante operação normal e sensibilidade reduzida ao ataque. Essa degradação de desempenho decorre da presença de dinâmicas não modeladas, ruídos de medição não-gaussianos e incertezas paramétricas inerentes ao sistema físico real, que não são completamente capturadas pelo modelo matemático utilizado no filtro de *Kalman*. Por outro lado, o método baseado na função de autocorrelação das inovações demonstrou maior robustez, detectando o ataque por ser menos dependente da precisão absoluta do modelo e das hipóteses sobre a distribuição do ruído. Esses resultados mostram que a transição de detectores projetados em ambiente simulado para aplicações práticas requer validação experimental e, eventualmente, ajustes nos parâmetros de detecção ou na escolha do método mais adequado às características do sistema físico.

Detecção por métodos alternativos: Embora o ataque seja furtivo para detectores baseados em resíduos operando na frequência do controlador, métodos de monitoramento que observem diretamente o comportamento do atuador (como sensores de corrente ou acelerômetros) poderiam identificar as oscilações de alta frequência características do ataque. Essa limitação sugere que a defesa em profundidade, combinando múltiplas camadas de monitoramento, pode ser eficaz contra essa classe de ataques.

Complexidade computacional para sistemas de grande porte: A otimização do sinal de ataque em sistemas MIMO de alta dimensão pode se tornar computacionalmente intensiva, especialmente quando combinada com altos fatores de superamostragem.

Embora o algoritmo GRASP tenha demonstrado bom desempenho nos casos estudados, a escalabilidade para sistemas industriais de grande porte requer investigação adicional.

Extensão para sistemas não lineares: As abordagens *Takagi-Sugeno* e *Koopman*, embora promissoras, introduzem erros de aproximação que se manifestam como erros de coincidência não nulos. A biblioteca de funções para o operador de *Koopman* requer conhecimento prévio sobre a estrutura da não linearidade, e a seleção inadequada pode comprometer a eficácia do ataque.

5.4 Perspectivas para Pesquisas Futuras

Os resultados obtidos abrem diversas direções para investigações futuras:

Desenvolvimento de contramedidas: A formulação do ataque sugere naturalmente estratégias de defesa que exploram as mesmas características temporais utilizadas pelo adversário. O projeto de controladores com taxa de amostragem variável ou arbitrária poderia dificultar a sincronização necessária para o ataque, uma vez que a condição de coincidência depende do conhecimento preciso dos instantes de amostragem; a introdução de *jitter* controlado no período de amostragem, mantido em segredo pelo sistema de controle, invalidaria o cálculo do sinal de ataque ótimo e forçaria erros de coincidência detectáveis. Esquemas de *watermarking* adaptados para detectar manipulações entre amostras representam outra direção promissora: ao invés de inserir sinais de autenticação apenas nos instantes de amostragem do controlador, a injeção de *watermarks* em frequências superiores permitiria verificar a integridade do canal de atuação nos intervalos inter-amostrais. A investigação de arquiteturas de controle intrinsecamente resilientes a ataques temporais constitui um campo de pesquisa relevante; controladores baseados em eventos (*event-triggered control*), que transmitem informações apenas quando certas condições são satisfeitas, eliminam a periodicidade previsível explorada pelo ataque. Outras contramedidas incluem: o uso de observadores de ordem reduzida operando em múltiplas taxas para detecção distribuída, a implementação de redundância temporal por meio de sensores com períodos de amostragem coprimos (ou seja, períodos cujo máximo divisor comum é unitário, como $T_1 = 7$ ms e $T_2 = 11$ ms, fazendo com que os instantes de amostragem simultânea sejam espaçados e irregulares, dificultando o projeto de um sinal de ataque que satisfaça a condição de coincidência para todos os sensores), e o desenvolvimento de controladores com realimentação contínua da integral do erro entre amostras, que penalizaria naturalmente desvios inter-amostrais mesmo sem detecção explícita do ataque.

Aplicações alternativas do Filtro de *Kalman* superamostrado: A estrutura do filtro operando em frequência superior à do controlador, desenvolvida neste trabalho para detecção de ataques, apresenta potencial para outras aplicações em sistemas de controle. Uma direção em investigação é a avaliação da qualidade de controladores *dead-beat*, onde

o monitoramento superamostrado permite analisar o comportamento entre amostras que permanece oculto na observação convencional, quantificando o *ripple* inter-amostral utilizando índices de degradação. Outras aplicações promissoras incluem: a detecção de taxa de amostragem inadequada, utilizando a função de autocorrelação do erro de estimação para revelar *aliasing* ou dinâmicas sub-amostradas; a identificação de dinâmicas não modeladas, onde picos na densidade espectral de potência do erro indicam modos não capturados pelo modelo de projeto; a avaliação de robustez a incertezas paramétricas, quantificando a sensibilidade do controlador por meio da divergência da covariância do erro; o monitoramento de sistemas *multi-rate*, verificando consistência quando sensores e atuadores operam em taxas diferentes; e a detecção de *jitter* em atrasos de comunicação, identificando variações no atraso por meio do erro de inovação nos instantes de transição. Essa linha de pesquisa está em andamento, com resultados preliminares em fase de produção de artigo científico.

Extensão para sistemas em rede: A formulação atual considera um único laço de controle. A extensão para sistemas de controle distribuído, onde múltiplos controladores interagem utilizando redes de comunicação, apresenta desafios adicionais relacionados à sincronização e à coordenação de ataques simultâneos. Em arquiteturas SCADA, onde RTUs comunicam-se com um centro de controle central por meio de protocolos como *Modbus* ou *DNP3*, o atacante poderia explorar os diferentes períodos de amostragem de cada RTU para planejar ataques de superamostragem coordenados que amplifiquem o impacto global enquanto mantêm furtividade local em cada subsistema. Em redes elétricas inteligentes (*smart grids*), a presença de múltiplos inversores, controladores de frequência e sistemas de proteção operando em escalas temporais distintas cria oportunidades para ataques que explorem as interações dinâmicas entre subsistemas. A formulação matemática para esses cenários requer a consideração de grafos de comunicação, atrasos variáveis entre nós, e a propagação de efeitos do ataque por meio da topologia da rede. Questões em aberto incluem: a otimização conjunta de sinais de ataque em múltiplos pontos de acesso, a análise de detectabilidade em arquiteturas de monitoramento hierárquico, e o desenvolvimento de métricas de impacto que capturem efeitos cascata em sistemas interconectados.

Aprendizado de máquina para detecção: Técnicas de aprendizado de máquina, especialmente redes neurais recorrentes e *autoencoders*, poderiam ser treinadas para identificar padrões anômalos característicos do ataque de superamostragem. Redes LSTM são particularmente adequadas para essa aplicação devido à sua capacidade de capturar dependências temporais de longo prazo nas sequências de resíduos do filtro de *Kalman*; treinadas com dados de operação normal, essas redes poderiam aprender a dinâmica esperada do sistema e sinalizar desvios sutis que precedem ou acompanham o ataque. *Autoencoders* variacionais oferecem outra abordagem promissora: ao aprender uma representação latente compacta do comportamento normal do sistema, anomalias se manifestariam como erros de reconstrução elevados, permitindo detecção mesmo quando o

ataque é projetado para evadir detectores estatísticos convencionais. A combinação dessas técnicas com o detector χ^2 proposto poderia resultar em sistemas de detecção híbridos mais robustos, onde o teste estatístico fornece garantias teóricas de desempenho sob hipóteses conhecidas, enquanto o componente de aprendizado de máquina captura padrões não modelados e adapta-se a variações nas características do ataque.

Análise de impacto econômico: A quantificação do custo do ataque em termos da redução na vida útil de componentes, consumo energético adicional e custos de manutenção permitiria uma análise de risco mais completa. O incremento na variação total do sinal de controle, observado experimentalmente neste trabalho (superiores a 7000%), traduz-se diretamente em ciclos de fadiga acelerados para atuadores como válvulas, motores e inversores. Modelos de fadiga baseados na regra de *Palmgren-Miner* (MINER, 1945) poderiam ser integrados à formulação do ataque para estimar a redução percentual da vida útil em função do fator de superamostragem e da duração do ataque. O consumo energético adicional, decorrente das oscilações de alta frequência impostas aos atuadores, representa outro vetor de custo mensurável utilizando a energia RMS do sinal de controle. Em ambientes industriais com tarifação por demanda de pico, essas oscilações podem resultar em penalidades contratuais significativas. Os custos de manutenção corretiva e preventiva também são afetados: falhas prematuras de componentes geram paradas não programadas, enquanto a detecção de desgaste anômalo pode exigir inspeções mais frequentes. A consolidação dessas métricas em um modelo de custo total de propriedade (TCO, do inglês *total cost of ownership*) sob ataque permitiria comparar quantitativamente o investimento em contramedidas (monitoramento superamostrado, redundância de sensores, criptografia de comunicações) com as perdas esperadas em cenários de ataque, fornecendo subsídios objetivos para decisões de gestão de risco em infraestruturas críticas.

Validação em sistemas industriais e extensão para sistemas variantes no tempo: Embora a planta *LAB-Volt* represente adequadamente dinâmicas de processos industriais, a validação em sistemas de produção, com todas as complexidades associadas, forneceria evidências mais claras sobre a aplicabilidade prática das formulações desenvolvidas. Ambientes industriais apresentam características não contempladas em bancadas didáticas: ruído de medição com distribuições não gaussianas, perturbações externas correlacionadas com ciclos de produção, variações de carga dependentes da demanda, envelhecimento de componentes que altera parâmetros do sistema, e interferências eletromagnéticas que afetam a comunicação. Uma extensão teórica necessária para aproximar a formulação da realidade industrial é a consideração de sistemas variantes no tempo, onde as matrizes (A, B, C) dependem de parâmetros operacionais como desgaste mecânico. Nesses casos, o ataque precisaria incorporar estimação *online* dos parâmetros da planta ou utilizar formulações robustas que garantam furtividade para uma família de modelos dentro de uma região de incerteza. Uma possibilidade para viabilizar tal robustez é o uso de zonótopos para representar os conjuntos de incerteza e limitar o espaço de estados do

ataque.

5.5 Considerações Finais

Essa dissertação evidenciou uma vulnerabilidade estrutural presente em todos os sistemas de controle digital: a existência de intervalos temporais entre amostras onde a dinâmica do sistema evolui sem observação direta. O ataque de superamostragem explora essa característica inerente ao controle amostrado para comprometer a integridade física do sistema enquanto mantém aparência de normalidade para os mecanismos de monitoramento convencionais.

A relevância deste trabalho reside não apenas na demonstração de uma nova classe de ameaça, mas principalmente na mudança de perspectiva que propõe para o projeto de sistemas de controle seguros. A segurança temporal, entendida como a garantia de comportamento seguro nos intervalos entre amostras, deve ser incorporada como requisito de projeto desde as fases iniciais de desenvolvimento, ao lado dos requisitos tradicionais de estabilidade e desempenho.

Os resultados experimentais confirmaram tanto a viabilidade prática do ataque quanto a eficácia do método de detecção proposto. O incremento de mais de 7000% na variação total do sinal de controle observado nos experimentos físicos ilustra o potencial destrutivo dessa classe de ataques, que poderia resultar em falhas prematuras de atuadores, aumento de custos operacionais e, em casos extremos, comprometimento da segurança de processos críticos.

Em um cenário onde sistemas ciberfísicos assumem papel cada vez mais central em infraestruturas críticas, a compreensão dessas vulnerabilidades e o desenvolvimento de contramedidas eficazes constituem imperativos para a engenharia de sistemas de controle.

Referências

- AANONSEN, C. E. Stuxnet, revisited (again): Producing the strategic relevance of cyber operations. *Journal of Cyber Policy*, v. 10, n. 1, p. 68–84, 2025. Citado na página 19.
- ABRAMS, M.; WEISS, J. *Malicious Control System Cyber Security Attack Case Study – Maroochy Water Services, Australia*. McLean, VA, USA, 2008. 16 p. Relatório técnico. <https://apps.dtic.mil/sti/pdfs/AD1107275.pdf>. Citado na página 18.
- AGUIRRE, L. A. *Introdução à Identificação de Sistemas: técnicas lineares e não lineares: teoria e aplicação*. 4. ed. Belo Horizonte: Editora UFMG, 2015. 730 p. ISBN 978-85-423-0079-6. Citado na página 108.
- ALNAJIM, A. M.; HABIB, S.; ISLAM, M.; THWIN, S. M.; ALOTAIBI, F. A comprehensive survey of cybersecurity threats, attacks, and effective countermeasures in industrial internet of things. *Technologies*, v. 11, n. 6, p. 161, nov. 2023. ISSN 2227-7080. Citado na página 18.
- ALRSLANI, F. A.; ALOHALI, M. A.; ALJEBREEN, M.; ALQAHTANI, H.; ALSHUHAIL, A.; ALSHAMMERI, M.; ALMUKADI, W. S. Enhancing cybersecurity via attribute reduction with deep learning model for false data injection attack recognition. *Scientific Reports*, v. 15, n. 1, p. 3944, 2025. Citado na página 32.
- AMIN, S.; LITRICO, X.; SASTRY, S.; BAYEN, A. M. Cyber security of water SCADA systems—part i: Analysis and experimentation of stealthy deception attacks. *IEEE Transactions on Control Systems Technology*, v. 21, n. 5, p. 1963–1970, 2012. Citado na página 42.
- BAHETI, R.; GILL, H. Cyber-physical systems. *The Impact of Control Technology*, IEEE Control Systems Society, v. 12, n. 1, p. 161–166, mar. 2011. Citado na página 17.
- BIANCHIN, G.; PASQUALETTI, F. Time-delay attacks in network systems. In: *Cyber-Physical Systems Security*. 2018. p. 157–174. Citado 2 vezes nas páginas 38 e 40.
- BOGGS, P. T.; TOLLE, J. W. Sequential quadratic programming. *Acta numerica*, v. 4, p. 1–51, 1995. Citado na página 51.
- BOYD, S.; VANDENBERGHE, L. *Convex optimization*. Cambridge university press, 2004. Citado na página 61.
- BROCKWELL, P. J.; DAVIS, R. A. *Introduction to time series and forecasting*. Springer, 2002. Citado na página 56.
- BRUMBACK, B.; SRINATH, M. A chi-square test for fault-detection in Kalman filters. *IEEE Transactions on Automatic Control*, v. 32, n. 6, p. 552–554, 1987. Citado na página 55.
- BRUNTON, S. L.; KUTZ, J. N. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge: Cambridge University Press, 2019. Citado na página 64.

- BUCY, R. S.; JOSEPH, P. D. *Filtering for stochastic processes with applications to guidance*. American Mathematical Soc., 2005. v. 326. Citado na página 54.
- CECILIA, A.; ASTOLFI, D.; CASADEI, G.; COSTA-CASTELLÓ, R.; NEŠIĆ, D. A masking protocol for private communication and attack detection in nonlinear observers. *2023 62nd IEEE Conference on Decision and Control (CDC)*. 2023. p. 7495–7500. Citado na página 46.
- CÉSAR, G. F.; CAMPOS, V. C. d. S.; AGUIRRE, L. A. Ataque de superamostragem. *Artigo aceito para publicação em congresso*, 2025. Simpósio Brasileiro de Automação Inteligente (SBAI 2025). Citado 2 vezes nas páginas 23 e 24.
- CHEN, T.; CECILIA, A.; ASTOLFI, D.; WANG, L.; LIU, Z.; SU, H. Chaotic masking protocol for secure communication and attack detection in remote estimation of cyber-physical systems. *IFAC-PapersOnLine*, v. 58, n. 21, p. 79–84, 2024. Citado na página 46.
- CRASSIDIS, J. L.; JUNKINS, J. L. *Optimal estimation of dynamic systems*. Chapman and Hall/CRC, 2004. Citado na página 54.
- DOKKU, N. S.; RAJ, R. D. A.; BODAPATI, S. K.; PALLAKONDA, A.; REDDY, Y. R. M.; PRAKASHA, K. K. Resilient cybersecurity in smart grid ics communication using blake3-driven dynamic key rotation and intrusion detection. *Scientific Reports*, v. 15, n. 1, p. 32754, 2025. Citado na página 32.
- DRAGOS. *ELECTRUM Targeted Ukrainian Electric Entity Using Custom Tools and CaddyWiper Malware, October 2022*. 2023. Publicação técnica. <https://www.dragos.com/blog/new-details-electrum-ukraine-electric-sector-compromise-2022>. Citado na página 19.
- ELLIOTT, J. NASA's advanced control law program for the f-8 digital fly-by-wire aircraft. *IEEE Transactions on Automatic Control*, v. 22, n. 5, p. 753–757, 1977. Citado 3 vezes nas páginas 26, 83 e 84.
- FANG, C.; QI, Y.; CHENG, P.; ZHENG, W. X. Optimal periodic watermarking schedule for replay attack detection in cyber-physical systems. *Automatica*, v. 112, p. 108698, 2020. Citado na página 35.
- FEO, T. A.; RESENDE, M. G. Greedy randomized adaptive search procedures. *Journal of global optimization*, v. 6, n. 2, p. 109–133, 1995. Citado na página 61.
- FESTO DIDACTIC. *Mobile Instrumentation and Process Control Training Systems, LabVolt Series 3500*. Eatontown, NJ, 2015. Datasheet. Citado 3 vezes nas páginas 25, 101 e 102.
- FITZGERALD, A. E.; KINGSLEY, C.; UMANS, S. D. *Electric Machinery*. 6. ed. New York, NY: McGraw-Hill, 2003. Citado na página 88.
- FRANKLIN, G. F.; POWELL, J. D.; WORKMAN, M. L. et al. *Digital control of dynamic systems*. Addison-wesley Menlo Park, 1998. v. 3. Citado na página 70.
- FRANZÈ, G.; TEDESCO, F.; FAMULARO, D. Resilience against replay attacks: A distributed model predictive control scheme for networked multi-agent systems. *IEEE/CAA Journal of Automatica Sinica*, v. 8, n. 3, p. 628–640, 2020. Citado na página 37.

- GANESH, P.; LOU, X.; CHEN, Y.; TAN, R.; YAU, D. K.; CHEN, D.; WINSLETT, M. Learning-based simultaneous detection and characterization of time delay attack in cyber-physical systems. *IEEE Transactions on Smart Grid*, v. 12, n. 4, p. 3581–3593, 2021. Citado na página 41.
- GILL, P. E.; MURRAY, W.; SAUNDERS, M. A. Snopt: An sqp algorithm for large-scale constrained optimization. *SIAM Review*, v. 47, n. 1, p. 99–131, 2005. Citado na página 51.
- GOLDENBERG, N.; WOOL, A. Accurate modeling of modbus/tcp for intrusion detection in SCADA systems. *International Journal of Critical Infrastructure Protection*, v. 6, n. 2, p. 63–75, 2013. Citado na página 45.
- GUO, H.; PANG, Z.; SUN, J.; LI, J. Detection of stealthy false data injection attacks against cyber-physical systems: A stochastic coding scheme. *Journal of Systems Science and Complexity*, v. 35, p. 1668–1684, 2022. Citado 3 vezes nas páginas 20, 21 e 22.
- GUO, H.; PANG, Z.-H.; SUN, J.; LI, J. An output-coding-based detection scheme against replay attacks in cyber-physical systems. *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 68, n. 10, p. 3306–3310, 2021. Citado na página 37.
- HARKAT, H.; CAMARINHA-MATOS, L. M.; GOES, J.; AHMED, H. F. T. Cyber-physical systems security: A systematic review. *Computers & Industrial Engineering*, v. 188, p. 109891, 2024. ISSN 0360-8352. Citado na página 18.
- HARRIS, T. J.; SEPPALA, C.; DESBOROUGH, L. A review of performance monitoring and assessment techniques for univariate and multivariate control systems. *Journal of process control*, v. 9, n. 1, p. 1–17, 1999. Citado na página 52.
- HU, P.; GAO, W.; LI, Y.; WU, M.; HUA, F.; QIAO, L. Detection of false data injection attacks in smart grids based on expectation maximization. *Sensors*, v. 23, n. 3, p. 1683, 2023. Citado na página 32.
- HUSSAIN, S. M. S.; USTUN, T. S.; KALAM, A. A review of IEC 62351 security mechanisms for IEC 61850 message exchanges. *IEEE Transactions on Industrial Informatics*, v. 16, n. 9, p. 5643–5654, 2020. Citado na página 45.
- INGIMUNDARSON, A.; HÄGGLUND, T. Robust tuning procedures of dead-time compensating controllers. *Control Engineering Practice*, v. 9, n. 11, p. 1195–1208, 2001. Citado na página 42.
- ISERMANN, R. *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer Science & Business Media, 2005. Citado 2 vezes nas páginas 55 e 56.
- JOO, Y.; QU, Z.; NAMERIKAWA, T. Resilient control of cyber-physical system using non-linear encoding signal against system integrity attacks. *IEEE Transactions on Automatic Control*, v. 66, n. 9, p. 4334–4341, set. 2021. Citado 2 vezes nas páginas 20 e 22.
- KALMAN, R. E. A new approach to linear filtering and prediction problems. 1960. Citado na página 53.
- KAY, S. M. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs: Prentice Hall, 1993. v. 1. Citado na página 55.

- KIM, K.; CHOI, S.; KWON, H.; KIM, H.; LIU, Z.; SEO, H. Page—practical aes-gcm encryption for low-end microcontrollers. *Applied Sciences*, v. 10, n. 9, p. 3131, 2020. Citado na página 46.
- KIM, T. T.; POOR, H. V. Strategic protection against data injection attacks on power grids. *IEEE Transactions on Smart Grid*, v. 2, n. 2, p. 326–333, 2011. Citado na página 33.
- KOOPMAN, B. O. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences*, v. 17, n. 5, p. 315–318, 1931. Citado na página 63.
- KOSUT, O.; JIA, L.; THOMAS, R. J.; TONG, L. Malicious data attacks on the smart grid. *IEEE Transactions on Smart Grid*, v. 2, n. 4, p. 645–658, 2011. Citado 2 vezes nas páginas 29 e 31.
- KUTZ, J. N.; BRUNTON, S. L.; BRUNTON, B. W.; PROCTOR, J. L. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. Philadelphia, PA: SIAM, 2016. Citado na página 64.
- LANGNER, R. Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security & Privacy*, v. 9, n. 3, p. 49–51, maio 2011. Citado na página 19.
- LEE, E. A. Cyber physical systems: Design challenges. In: *2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*. Orlando, FL, 2008. p. 363–369. Citado na página 17.
- LIANG, G.; ZHAO, J.; LUO, F.; WELLER, S. R.; DONG, Z. Y. A review of false data injection attacks against modern power systems. *IEEE Transactions on Smart Grid*, v. 8, n. 4, p. 1630–1638, 2017. Citado 3 vezes nas páginas 21, 29 e 31.
- LIN, H.; SLAGELL, A.; MARTINO, C. D.; KALBARCZYK, Z.; IYER, R. K. Adapting bro into SCADA: building a specification-based intrusion detection system for the dnp3 protocol. In: *Proceedings of the Eighth Annual Cyber Security and Information Intelligence Research Workshop*. 2013. p. 1–4. Citado na página 45.
- LIU, Y.; NING, P.; REITER, M. K. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security (TISSEC)*, ACM New York, NY, USA, v. 14, n. 1, p. 1–33, 2011. Citado 2 vezes nas páginas 28 e 30.
- LJUNG, G. M.; BOX, G. E. On a measure of lack of fit in time series models. *Biometrika*, v. 65, n. 2, p. 297–303, 1978. Citado na página 56.
- LJUNG, L. *System Identification: Theory for the User*. Prentice Hall PTR, 1999. ISBN 9780136566953. Citado 3 vezes nas páginas 107, 108 e 109.
- LOU, X.; TRAN, C.; TAN, R.; YAU, D. K.; KALBARCZYK, Z. T. Assessing and mitigating impact of time delay attack: a case study for power grid frequency control. In: *Proceedings of the 10th ACM/IEEE international conference on cyber-physical systems*. 2019. p. 207–216. Citado 2 vezes nas páginas 40 e 41.
- MAGLARAS, L. A.; JIANG, J. Intrusion detection in SCADA systems using machine learning techniques. In: *2014 Science and Information Conference*. 2014. p. 626–631. Citado na página 46.

- MAYBECK, P. S. *Stochastic models, estimation, and control*. Academic press, 1982. v. 3. Citado na página 53.
- MEHRA, R. On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on automatic control*, v. 15, n. 2, p. 175–184, 1970. Citado na página 56.
- MEHRA, R. K.; PESCHON, J. An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica*, v. 7, n. 5, p. 637–640, 1971. Citado na página 21.
- MEZIĆ, I. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, v. 41, n. 1, p. 309–325, 2005. Citado na página 63.
- MINER, M. A. Cumulative damage in fatigue. American Society of Mechanical Engineers, 1945. Citado na página 127.
- MO, Y.; CHABUKSWAR, R.; SINOPOLI, B. Detecting integrity attacks on SCADA systems. *IEEE Transactions on Control Systems Technology*, v. 22, n. 4, p. 1396–1407, 2013. Citado 3 vezes nas páginas 28, 31 e 32.
- MO, Y.; SINOPOLI, B. Secure control against replay attacks. In: *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Monticello, IL, 2009. p. 911–918. Citado 5 vezes nas páginas 20, 21, 22, 33 e 35.
- MO, Y.; SINOPOLI, B. On the performance degradation of cyber-physical systems under stealthy integrity attacks. *IEEE Transactions on Automatic Control*, v. 61, n. 9, p. 2618–2624, set. 2016. Citado 2 vezes nas páginas 20 e 21.
- NANKYA, M.; CHATAUT, R.; AKL, R. Securing industrial control systems: Components, cyber threats, and machine learning-driven defense strategies. *Sensors*, Basel, Switzerland, v. 23, n. 21, p. 8840, out. 2023. ISSN 1424-8220. Citado 2 vezes nas páginas 17 e 18.
- NOCEDAL, J.; WRIGHT, S. J. *Numerical optimization*. Springer, 2006. Livro. Citado 2 vezes nas páginas 51 e 61.
- OGATA, K. *Discrete-time control systems*. Prentice-Hall, Inc., 1995. Citado na página 71.
- PASQUALETTI, F.; DÖRFLER, F.; BULLO, F. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, v. 58, n. 11, p. 2715–2729, nov. 2013. Citado 2 vezes nas páginas 20 e 21.
- PONOMAREV, S.; ATKISON, T. Industrial control system network intrusion detection by telemetry analysis. *IEEE Transactions on Dependable and Secure Computing*, v. 13, n. 2, p. 252–260, 2016. Citado na página 46.
- PROSKA, K.; WOLFRAM, J.; WILSON, J.; BLACK, D.; LUNDEN, K.; ZAFRA, D. K.; BRUBAKER, N.; MCLELLAN, T.; SISTRUNK, C. *Sandworm Disrupts Power in Ukraine Using a Novel Attack Against Operational Technology*. 2023. Publicação técnica (Google Cloud Threat Intelligence). <https://cloud.google.com/blog/topics/threat-intelligence/sandworm-disrupts-power-ukraine-operational-technology/>. Citado na página 19.
- QIN, S. J.; BADGWELL, T. A. A survey of industrial model predictive control technology. *Control Engineering Practice*, v. 11, n. 7, p. 733–764, 2003. Citado na página 52.

- REN, G.; LIU, H.; ZHANG, Y.; LI, Y. Attack detection based on encoding–decoding approach for cyber–physical systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, v. 53, n. 10, p. 6293–6302, out. 2023. Citado na página 21.
- ROCKAFELLAR, R. *Convex Analysis*. Princeton University Press, 1997. ISBN 9780691015866. Citado na página 61.
- RUBIO, J. E.; ALCARAZ, C.; ROMAN, R.; LOPEZ, J. Current cyber-defense trends in industrial control systems. *Computers & Security*, v. 87, p. 101561, 2019. ISSN 0167-4048. Citado 3 vezes nas páginas 17, 18 e 19.
- SAMTANI, S.; YU, S.; ZHU, H.; PATTON, M.; CHEN, H. Identifying SCADA vulnerabilities using passive and active vulnerability assessment techniques. In: *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*. 2016. p. 25–30. Citado 2 vezes nas páginas 43 e 44.
- SANCHEZ, H. S.; ROTONDO, D.; ESCOBET, T.; PUIG, V.; SALUDES, J.; QUEVEDO, J. Detection of replay attacks in cyber-physical systems using a frequency-based signature. *Journal of the Franklin Institute*, v. 356, n. 5, p. 2798–2824, 2019. Citado 2 vezes nas páginas 35 e 36.
- SANISLAV, T.; MICLEA, L. Cyber-physical systems – concept, challenges and research areas. *Control Engineering and Applied Informatics*, v. 14, n. 2, p. 28–33, 2012. Citado na página 17.
- SARGOLZAEI, A.; YEN, K. K.; ABDELGHANI, M. N. Preventing time-delay switch attack on load frequency control in distributed power systems. *IEEE Transactions on Smart Grid*, v. 7, n. 2, p. 1176–1185, 2015. Citado 2 vezes nas páginas 38 e 42.
- SARGOLZAEI, A.; YEN, K. K.; ABDELGHANI, M. N.; SARGOLZAEI, S.; CARBUNAR, B. Resilient design of networked control systems under time delay switch attacks, application in smart grid. *IEEE Access*, v. 5, p. 15901–15912, 2017. Citado 2 vezes nas páginas 40 e 42.
- SATCHIDANANDAN, B.; KUMAR, P. R. Dynamic watermarking: Active defense of networked cyber–physical systems. *Proceedings of the IEEE*, v. 105, n. 2, p. 219–240, fev. 2017. Citado na página 22.
- SAYGHE, A. Digital twin-driven intrusion detection for industrial SCADA: A cyber-physical case study. *Sensors*, v. 25, n. 16, p. 4963, 2025. Citado na página 32.
- SCHLEGEL, R.; OBERMEIER, S.; SCHNEIDER, J. A security evaluation of IEC 62351. *Journal of Information Security and Applications*, v. 34, p. 197–204, 2017. Citado na página 45.
- SHARMA, R.; JOSHI, A. M.; SAHU, C.; NANDA, S. J. Detection of false data injection in smart grid using PCA based unsupervised learning. *Electrical Engineering*, v. 105, n. 4, p. 2383–2396, 2023. Citado na página 32.
- SHEREEN, E.; BITARD, F.; DÁN, G.; SEL, T.; FRIES, S. Next steps in security for time synchronization: Experiences from implementing IEEE 1588 v2.1. In: *2019 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication (ISPCS)*. 2019. p. 1–6. Citado na página 45.

- SKOGESTAD, S.; POSTLETHWAITE, I. *Multivariable feedback control: analysis and design*. John Wiley & Sons, 2005. Citado na página 52.
- TAKAGI, T.; SUGENO, M. Fuzzy identification of systems and its applications to modeling and control. *IEEE transactions on systems, man, and cybernetics*, n. 1, p. 116–132, 1985. Citado na página 57.
- TANAKA, K.; WANG, H. O. *Fuzzy control systems design and analysis: a linear matrix inequality approach*. John Wiley & Sons, 2004. Citado na página 57.
- TEIXEIRA, A.; SOU, K. C.; SANDBERG, H.; JOHANSSON, K. H. Secure control systems: A quantitative risk management approach. *IEEE Control Systems Magazine*, v. 35, n. 1, p. 24–45, 2015. Citado na página 46.
- URBINA, D. I.; GIRALDO, J. A.; CARDENAS, A. A.; TIPPENHAUER, N. O.; VALENTE, J.; FAISAL, M.; RUTHS, J.; CANDELL, R.; SANDBERG, H. Limiting the impact of stealthy attacks on industrial control systems. In: *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 2016. p. 1092–1105. Citado 2 vezes nas páginas 31 e 33.
- WILLIAMS, M. O.; KEVREKIDIS, I. G.; ROWLEY, C. W. A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, v. 25, n. 6, p. 1307–1346, 2015. Citado na página 63.
- XIAHOU, K. S.; LIU, Y.; WU, Q. H. Robust load frequency control of power systems against random time-delay attacks. *IEEE Transactions on Smart Grid*, v. 12, n. 1, p. 909–911, jan. 2021. Citado 2 vezes nas páginas 21 e 22.
- YANG, W.; ZHENG, Z.; CHEN, G.; TANG, Y.; WANG, X. Security analysis of a distributed networked system under eavesdropping attacks. *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 67, n. 7, p. 1254–1258, jul. 2020. Citado na página 20.
- YANG, Y.; MCLAUGHLIN, K.; LITTLER, T.; SEZER, S.; PRANGGONO, B.; WANG, H. F. Intrusion detection system for IEC 60870-5-104 based SCADA networks. In: *2013 IEEE Power & Energy Society General Meeting*. 2013. p. 1–5. Citado na página 45.
- YANG, Y.; MCLAUGHLIN, K.; SEZER, S.; LITTLER, T.; IM, E. G.; PRANGGONO, B.; WANG, H. Multiattribute SCADA-specific intrusion detection system for power networks. *IEEE Transactions on Power Delivery*, v. 29, n. 3, p. 1092–1102, 2014. Citado na página 45.
- ZDZIKOT, T. Cyberspace and cybersecurity. In: CHAŁUBIŃSKA-JENTKIEWICZ, K.; RADONIEWICZ, F.; ZIELIŃSKI, T. (Ed.). *Cybersecurity in Poland*. Cham: Springer, 2021. p. 9–21. ISBN 978-3-030-78550-5. Citado na página 18.
- ZHANG, G.; LI, J.; BAMISILE, O.; CAI, D.; HUANG, Q. A novel data-driven time-delay attack evaluation method for wide-area cyber-physical smart grid systems. *Sustainable Energy, Grids and Networks*, v. 32, p. 100960, 2022. Citado na página 41.
- ZHOU, M.; ZHANG, Z.; XIE, L. Permutation entropy based detection scheme of replay attacks in industrial cyber-physical systems. *Journal of the Franklin Institute*, v. 358, n. 7, p. 4058–4076, 2021. Citado na página 36.

ZHU, M.; MARTINEZ, S. On the performance analysis of resilient networked control systems under replay attacks. *IEEE Transactions on Automatic Control*, v. 59, n. 3, p. 804–808, 2013. Citado 2 vezes nas páginas 33 e 36.

Apêndices

APÊNDICE A – Algoritmos Implementados

A.1 Ataque de Superamostragem T-S MIMO via GRASP (C-GRASP)

Algorithm 1 – Ataque de Superamostragem T-S MIMO via GRASP (C-GRASP)

Require: Sistema T-S $\{(\mathbf{A}_i, \mathbf{B}_i)\}_{i=1}^{r_{\text{regras}}}$, $\mathbf{x}_k$, $\alpha \in [0,1]$, limites $[\mathbf{u}, \bar{\mathbf{u}}]$

Ensure: Sequência de ataque $\mathbf{u}_a^* \in \mathbb{R}^{rm}$

- 1: **Fase Offline:**
- 2: Discretizar matrizes T-S com T_d e T_D : obter $\{\mathbf{A}_{D,i}, \mathbf{B}_{D,i}, \bar{\mathbf{A}}_i, \bar{\mathbf{B}}_i\}$
- 3: Projetar ganhos PDC $\{\mathbf{K}_i\}_{i=1}^{r_{\text{regras}}}$, em que $\mathbf{K}_i \in \mathbb{R}^{m \times n}$
- 4: **Fase Online (a cada kT_D):**
- 5: Medir $\mathbf{x}_k$ e calcular $h_{i,k}$ via Equação (3.37)
- 6: Calcular $\bar{\mathbf{A}}_d(h_k)$, $\bar{\mathbf{B}}_d(h_k)$, $\mathbf{A}_D(h_k)$, $\mathbf{B}_D(h_k)$
- 7: Construir $\mathcal{C}_r(h_k) \in \mathbb{R}^{n \times rm}$ via Equação (3.54)
- 8: Calcular $\mathbf{d}_k(h_k) \in \mathbb{R}^n$ via Equação (3.55)
- 9: **if** $\text{rank}(\mathcal{C}_r) < n$ **then**
- 10: **return** Falha (sistema não controlável)
- 11: **end if**
- 12: **Preparação da Otimização:**
- 13: $\mathbf{u}_a^{(0)} \leftarrow \mathcal{C}_r^\top (\mathcal{C}_r \mathcal{C}_r^\top)^{-1} \mathbf{d}_k \in \mathbb{R}^{rm}$
- 14: $\mathbf{N}_r \leftarrow \text{null}(\mathcal{C}_r) \in \mathbb{R}^{rm \times (rm-n)}$
- 15: $n_\eta \leftarrow rm - n$
- 16: **if** $n_\eta = 0$ **then**
- 17: $\mathbf{u}_a^* \leftarrow \mathbf{u}_a^{(0)}$
- 18: **else**
- 19: $J_{\max} \leftarrow -\infty$, $\mathbf{u}_a^* \leftarrow \mathbf{u}_a^{(0)}$
- 20: **for** $i = 1$ to $N_{\text{tentativas}}$ **do**
- 21: $\boldsymbol{\eta}_0 \leftarrow \mathbf{0}_{n_\eta}$
- 22: // Fase 1: Construção Gulosa Randomizada (GRASP)
- 23: **for** $s = 1$ to n_η **do**
- 24: Definir grade de candidatos discretos $\mathcal{D}_s$ para a componente s
- 25: Avaliar custo incremental $g(\nu)$ para cada $\nu \in \mathcal{D}_s$
- 26: $g_{\max} \leftarrow \max(g)$, $g_{\min} \leftarrow \min(g)$
- 27: $\text{RCL} \leftarrow \{\nu \in \mathcal{D}_s \mid g(\nu) \geq g_{\max} - \alpha(g_{\max} - g_{\min})\}$

```

28:          $[\boldsymbol{\eta}_0]_s \leftarrow \text{Random}(\text{RCL})$ 
29:     end for
30:     // Fase 2: Busca Local (Refinamento Contínuo)
31:      $\boldsymbol{\eta}_i^* \leftarrow \text{fmincon}(-J, \boldsymbol{\eta}_0, \text{restrições de saturação})$ 
32:      $\mathbf{u}_i \leftarrow \mathbf{u}_a^{(0)} + \mathbf{N}_r \boldsymbol{\eta}_i^*$ 
33:     if  $J(\mathbf{u}_i) > J_{\max}$  then
34:          $J_{\max} \leftarrow J(\mathbf{u}_i), \mathbf{u}_a^* \leftarrow \mathbf{u}_i$ 
35:     end if
36: end for
37: end if
38: Aplicar Ataque:
39: for  $j = 0$  to  $r - 1$  do
40:     Extrair  $\bar{\mathbf{u}}_{\ell+j} = [\mathbf{u}_a^*]_{jm+1:(j+1)m} \in \mathbb{R}^m$ 
41:     Aplicar  $\bar{\mathbf{u}}_{\ell+j}$  em  $t = kT_D + jT_d$ 
42: end for
43: return  $\mathbf{u}_a^*$ 

```

A.2 Ataque de Superamostragem MIMO via *Koopman* (EDMD-GRASP)

Algorithm 2 – Ataque de Superamostragem MIMO via *Koopman* (EDMD-GRASP)

Require: Trajetória de dados $\{(\mathbf{x}_i, \mathbf{u}_i, \mathbf{x}_{i+1})\}_{i=1}^M$, dicionário $\boldsymbol{\psi}(\cdot)$, fator r , limites $[\underline{\mathbf{u}}, \bar{\mathbf{u}}]$, $\alpha \in [0, 1]$, regularização $\lambda > 0$

Ensure: Sequência de ataque $\mathbf{u}_a^* \in \mathbb{R}^{rm}$

- 1: **Fase *Offline* (Identificação via EDMD):**
- 2: Aplicar dicionário aos dados: $\mathbf{Z} \leftarrow [\boldsymbol{\psi}(\mathbf{x}_1), \dots, \boldsymbol{\psi}(\mathbf{x}_M)] \in \mathbb{R}^{N \times M}$
- 3: $\mathbf{Z}' \leftarrow [\boldsymbol{\psi}(\mathbf{x}_2), \dots, \boldsymbol{\psi}(\mathbf{x}_{M+1})] \in \mathbb{R}^{N \times M}$
- 4: $\mathbf{U} \leftarrow [\mathbf{u}_1, \dots, \mathbf{u}_M] \in \mathbb{R}^{m \times M}$
- 5: Construir matriz de regressores: $\mathbf{R} \leftarrow [\mathbf{Z}^\top; \mathbf{U}^\top]^\top \in \mathbb{R}^{(N+m) \times M}$
- 6: Identificar matrizes de *Koopman* com regularização de *Tikhonov* via Equação (3.71):
- 7: $[\mathbf{A}_K \ \mathbf{B}_K] \leftarrow \mathbf{Z}' \mathbf{R}^\top (\mathbf{R} \mathbf{R}^\top + \lambda \mathbf{I})^{-1}$, em que $\mathbf{A}_K \in \mathbb{R}^{N \times N}$, $\mathbf{B}_K \in \mathbb{R}^{N \times m}$
- 8: Discretizar ou reidentificar com período T_d : obter $\bar{\mathbf{A}}_{d,K} \in \mathbb{R}^{N \times N}$, $\bar{\mathbf{B}}_{d,K} \in \mathbb{R}^{N \times m}$
- 9: Calcular matrizes agregadas: $\bar{\boldsymbol{\Theta}}_K \leftarrow [\bar{\mathbf{A}}_{d,K}]^r$ via Equação (3.78)
- 10: $\bar{\boldsymbol{\Gamma}}_K \leftarrow [[\bar{\mathbf{A}}_{d,K}]^{r-1} \bar{\mathbf{B}}_{d,K}, \dots, \bar{\mathbf{B}}_{d,K}] \in \mathbb{R}^{N \times rm}$ via Equação (3.79)

```

11: Construir matriz de projeção física  $\mathbf{C}_{phys} \in \mathbb{R}^{n \times N}$  via Equação (3.82)
12: Calcular matriz de controlabilidade efetiva:  $\mathbf{C}_r^K \leftarrow \mathbf{C}_{phys} \bar{\Gamma}_K \in \mathbb{R}^{n \times rm}$  via Equação (3.84)
13: Fase Online (a cada  $kT_D$ ):
14: Medir estado atual  $\mathbf{x}_k$  e interceptar controle legítimo  $\mathbf{u}_k^* \in \mathbb{R}^m$ 
15: Elevar estados:  $\mathbf{z}_k \leftarrow \psi(\mathbf{x}_k) \in \mathbb{R}^N$ 
16: Calcular estado nominal futuro elevado:  $\mathbf{z}_{k+1}^{nom} \leftarrow \mathbf{A}_K \mathbf{z}_k + \mathbf{B}_K \mathbf{u}_k^*$ 
17: Calcular vetor de desvio completo:  $\mathbf{d}_k^{K,full} \leftarrow \mathbf{z}_{k+1}^{nom} - \bar{\Theta}_K \mathbf{z}_k \in \mathbb{R}^N$ 
18: Projetar para espaço físico:  $\mathbf{d}_k^K \leftarrow \mathbf{C}_{phys} \mathbf{d}_k^{K,full} \in \mathbb{R}^n$  via Equação (3.85)
19: if  $\text{rank}(\mathbf{C}_r^K) < n$  then
20:     return Falha (sistema não controlável)
21: end if
22: Preparação da Otimização:
23:  $\mathbf{u}_a^{(0)} \leftarrow (\mathbf{C}_r^K)^\top (\mathbf{C}_r^K (\mathbf{C}_r^K)^\top)^{-1} \mathbf{d}_k^K \in \mathbb{R}^{rm}$ 
24:  $\mathbf{N}_r \leftarrow \text{null}(\mathbf{C}_r^K) \in \mathbb{R}^{rm \times (rm-n)}$ 
25:  $n_\eta \leftarrow rm - n$ 
26: if  $n_\eta = 0$  then
27:      $\mathbf{u}_a^* \leftarrow \mathbf{u}_a^{(0)}$ 
28: else
29:      $J_{\max} \leftarrow -\infty$ ,  $\mathbf{u}_a^* \leftarrow \mathbf{u}_a^{(0)}$ 
30:     for  $i = 1$  to  $N_{tentativas}$  do
31:          $\boldsymbol{\eta}_0 \leftarrow \mathbf{0}_{n_\eta}$ 
32:         // Fase 1: Construção Gulosa Randomizada (GRASP)
33:         for  $s = 1$  to  $n_\eta$  do
34:             Definir grade de candidatos discretos  $\mathcal{D}_s$  para a componente  $s$ 
35:             Avaliar custo incremental  $g(\nu)$  para cada  $\nu \in \mathcal{D}_s$ :
36:              $g(\nu) \leftarrow J(\mathbf{u}_a^{(0)} + \mathbf{N}_r[\boldsymbol{\eta}_0; \nu])$  em que  $[\cdot; \cdot]$  indica inserção em posição  $s$ 
37:              $g_{\max} \leftarrow \max_{\nu \in \mathcal{D}_s} g(\nu)$ ,  $g_{\min} \leftarrow \min_{\nu \in \mathcal{D}_s} g(\nu)$ 
38:              $\text{RCL} \leftarrow \{\nu \in \mathcal{D}_s \mid g(\nu) \geq g_{\max} - \alpha(g_{\max} - g_{\min})\}$ 
39:              $[\boldsymbol{\eta}_0]_s \leftarrow \text{Random}(\text{RCL})$ 
40:         end for
41:         // Fase 2: Busca Local (Refinamento Contínuo)
42:         Definir função objetivo:  $f(\boldsymbol{\eta}) = -\|\mathbf{u}_a^{(0)} + \mathbf{N}_r \boldsymbol{\eta}\|^2$ 
43:         Definir restrições:  $\underline{\mathbf{u}} \leq [\mathbf{u}_a^{(0)} + \mathbf{N}_r \boldsymbol{\eta}]_{(j-1)m+1:jm} \leq \bar{\mathbf{u}}$ ,  $j = 1, \dots, r$ 
44:          $\boldsymbol{\eta}_i^* \leftarrow \text{fmincon}(f, \boldsymbol{\eta}_0, \text{restrições})$ 
45:          $\mathbf{u}_i \leftarrow \mathbf{u}_a^{(0)} + \mathbf{N}_r \boldsymbol{\eta}_i^*$ 
46:         if  $J(\mathbf{u}_i) > J_{\max}$  then
47:              $J_{\max} \leftarrow J(\mathbf{u}_i)$ ,  $\mathbf{u}_a^* \leftarrow \mathbf{u}_i$ 
48:         end if
49:     end for

```

```

50: end if
51: Aplicar Ataque:
52: for  $j = 0$  to  $r - 1$  do
53:   Extrair  $\bar{\mathbf{u}}_{\ell+j} = [\mathbf{u}_a^*]_{jm+1:(j+1)m} \in \mathbb{R}^m$ 
54:   Aplicar  $\bar{\mathbf{u}}_{\ell+j}$  em  $t = kT_D + jT_d$ 
55: end for
56: return  $\mathbf{u}_a^*$ 

```

APÊNDICE B – Implementação Experimental de Ataques Cibernéticos Clássicos

Este apêndice apresenta os resultados experimentais e as análises de impacto referentes aos ataques cibernéticos clássicos: Injeção de Dados Falsos (FDI), *Replay* e Atraso Temporal. A fundamentação teórica para estes ataques encontra-se detalhada, respectivamente, nas Seções 2.1, 2.2 e 2.3 do texto principal.

B.1 Ataque de Injeção de Dados Falsos

O ataque de injeção de dados falsos consiste na modificação das medições ou comandos de controle para induzir comportamentos indesejados no sistema.

B.1.1 Injeção Aditiva (*Bias*)

O atacante adiciona um sinal de ataque $a(k)$ às medições legítimas:

$$y_a(k) = y(k) + a(k). \quad (\text{B.1})$$

Quando $a(k) = a_0$ é constante, denomina-se ataque de *bias*. Esse tipo de injeção faz com que o controlador perceba a saída como sendo maior do que realmente é, induzindo uma redução na ação de controle. Em regime permanente, o sistema converge para um novo ponto de operação deslocado do *setpoint* desejado.

A Figura 42 apresenta os resultados de simulação para diferentes valores de *bias*. Observa-se que o erro de regime permanente é aproximadamente igual ao valor do *bias* injetado, confirmando a relação:

$$y_{ss} \approx r - a_0, \quad (\text{B.2})$$

em que r é o *setpoint* e y_{ss} é a saída em regime permanente.

B.1.2 Injeção Multiplicativa

O atacante escala as medições por um fator α :

$$y_a(k) = \alpha \cdot y(k), \quad \alpha \neq 1. \quad (\text{B.3})$$

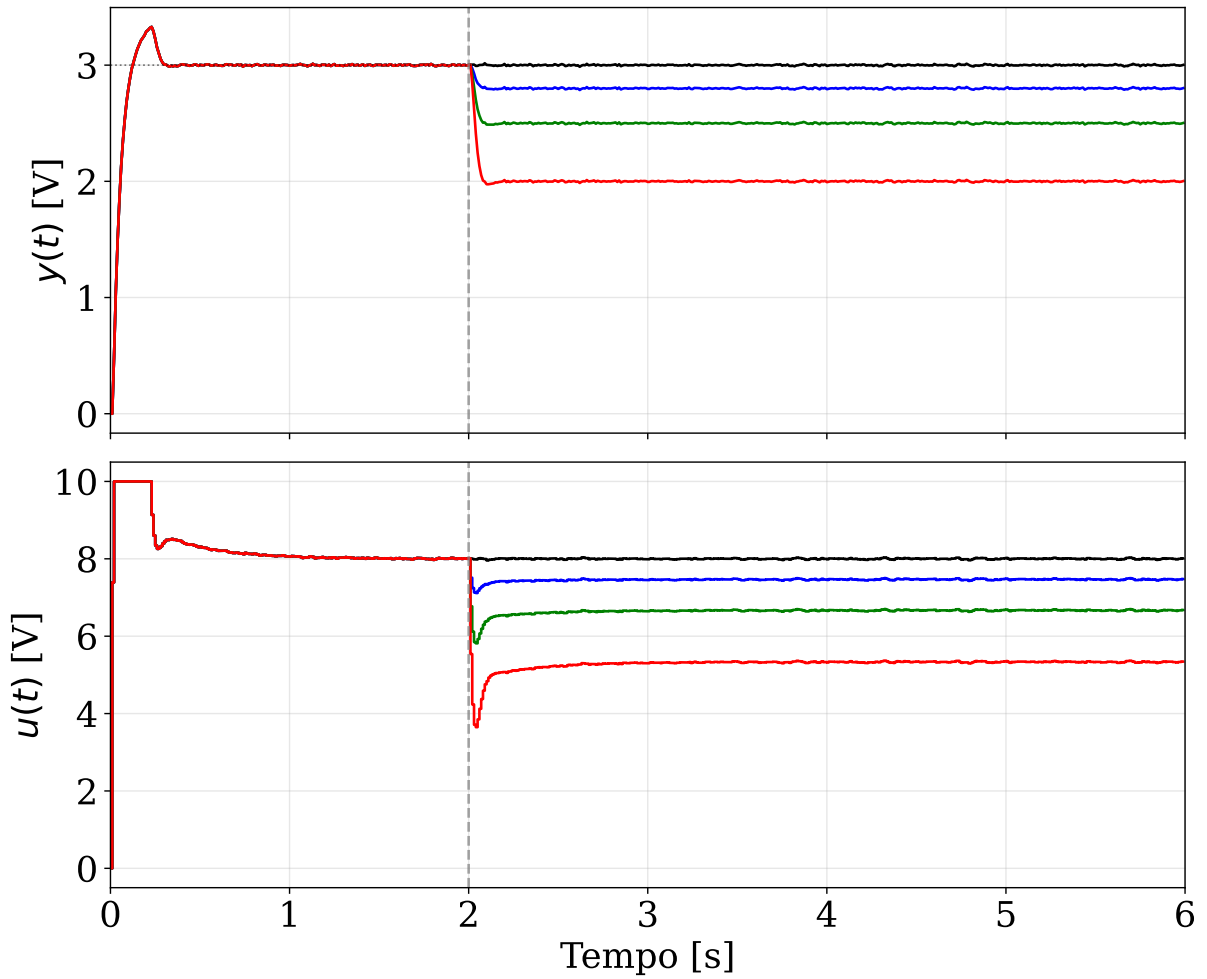


Figura 42 – Resposta do sistema sob ataque FDI aditivo (*bias*) iniciado em $t = 2$ s (linha vertical tracejada cinza): o painel superior mostra saída $y(t)$; e o painel inferior mostra a ação de controle $u(t)$. Preto: sem ataque ($a_0 = 0$); azul: $a_0 = 0,2$ V; verde: $a_0 = 0,5$ V; vermelho: $a_0 = 1,0$ V. A linha pontilhada indica o *setpoint* de 3 V.

Fatores $\alpha > 1$ fazem o controlador perceber uma saída maior que a real, reduzindo a ação de controle e causando queda na saída. Fatores $\alpha < 1$ têm efeito oposto, elevando a saída acima do *setpoint*. Em regime permanente, a relação entre a saída e o fator de escala é:

$$y_{ss} = \frac{r}{\alpha}. \quad (\text{B.4})$$

A Figura 43 ilustra esse comportamento. Para $\alpha = 0,8$, a saída converge para $y_{ss} \approx 3,75$ V, enquanto para $\alpha = 1,5$, converge para $y_{ss} \approx 2,0$ V. Nota-se que o ataque multiplicativo com $\alpha < 1$ pode levar a ação de controle à saturação, como observado para $\alpha = 0,8$.

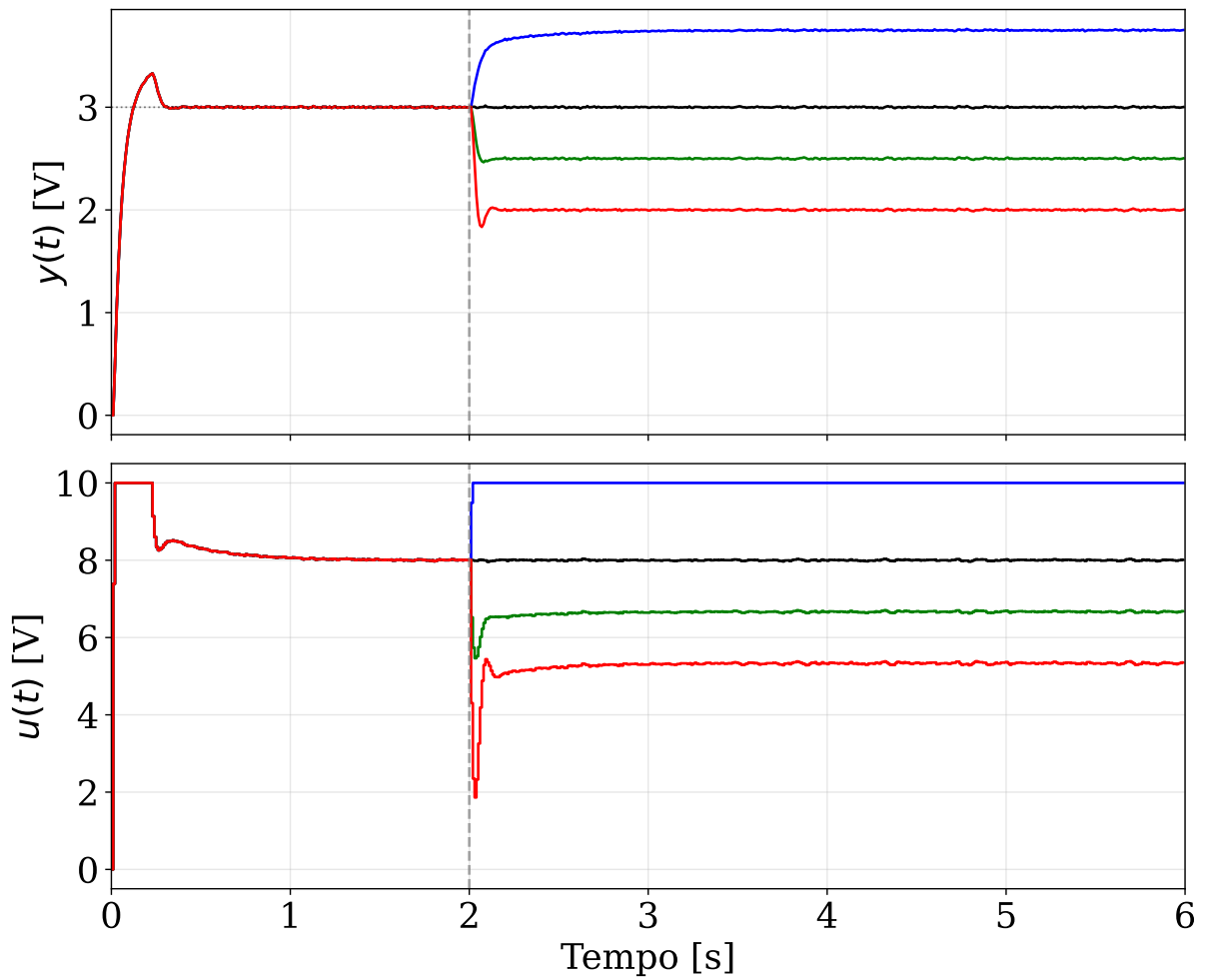


Figura 43 – Resposta do sistema sob ataque FDI multiplicativo iniciado em $t = 2$ s (linha vertical tracejada cinza): o painel superior mostra saída $y(t)$; e o painel inferior mostra a ação de controle $u(t)$. Preto: sem ataque ($\alpha = 1,0$); azul: $\alpha = 0,8$; verde: $\alpha = 1,2$; vermelho: $\alpha = 1,5$.

B.1.3 Injeção Senoidal

O atacante pode injetar um sinal periódico para induzir oscilações no sistema:

$$a(k) = A_a \sin(2\pi f_a k T_s), \quad (\text{B.5})$$

em que A_a é a amplitude e f_a é a frequência do sinal de ataque. Esse tipo de ataque pode ser utilizado para causar fadiga em componentes mecânicos ou para mascarar outras atividades maliciosas.

A Figura 44 apresenta os resultados para $A_a = 0,3$ V e $f_a = 2$ Hz. A saída real oscila em torno do *setpoint* com amplitude aproximadamente igual à amplitude do sinal de ataque, enquanto o controlador tenta compensar as variações percebidas.

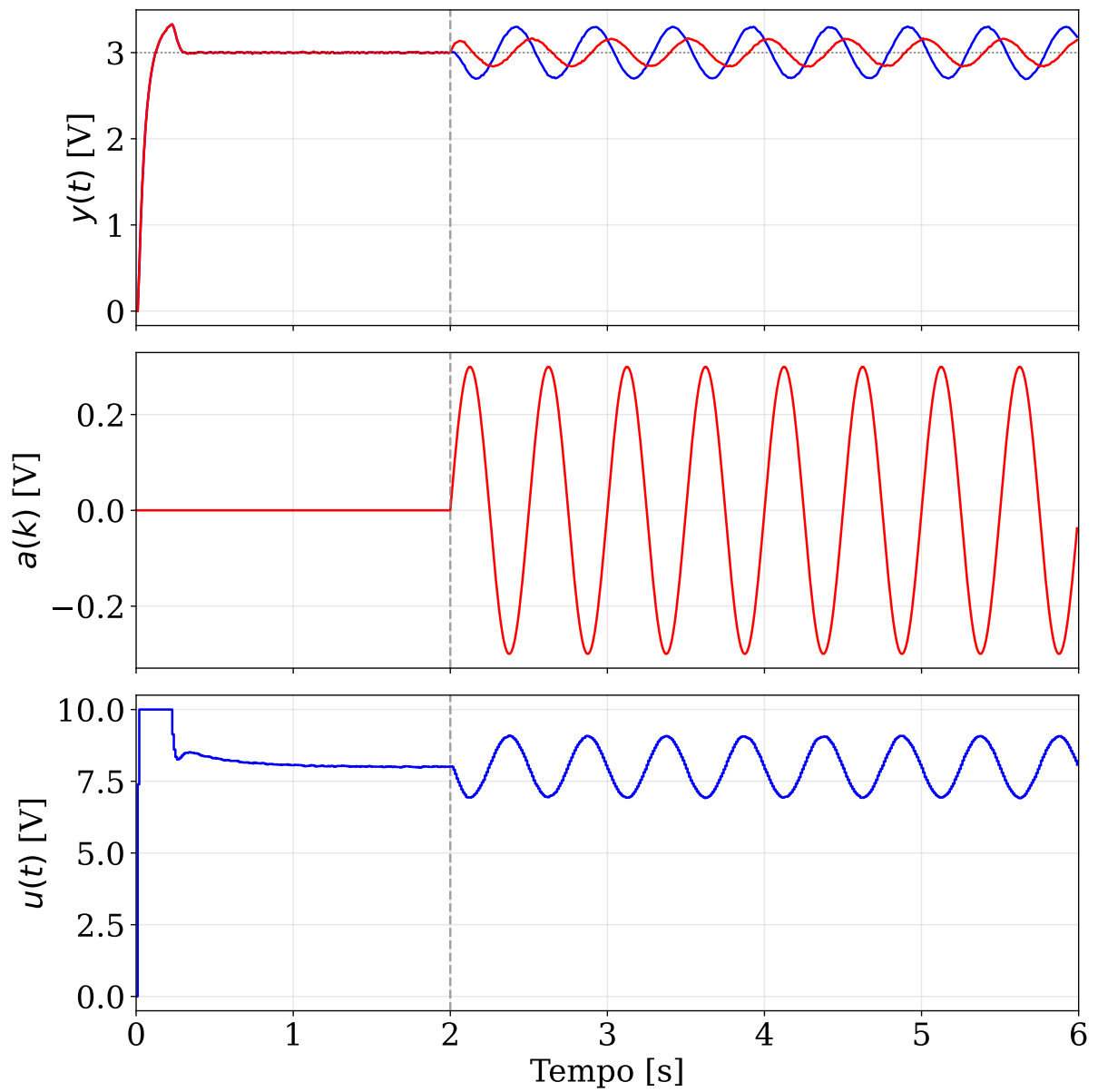


Figura 44 – Resposta do sistema sob ataque FDI senoidal iniciado em $t = 2$ s (linha vertical tracejada cinza) com amplitude $A_a = 0,3$ V e frequência $f_a = 2$ Hz: o painel superior mostra a saída do sistema, em que azul representa a saída real $y(t)$ e vermelho a medição atacada $y_a(t)$; o painel central mostra o sinal de ataque $a(k)$; e o painel inferior mostra a ação de controle $u(t)$.

B.1.4 Análise e Detecção

A Tabela 17 resume as métricas de desempenho do sistema sob os diferentes tipos de ataque FDI. O erro médio absoluto é calculado como $\bar{e} = \frac{1}{N} \sum_k |y(k) - r|$ durante a fase de ataque.

Tabela 17 – Métricas de desempenho sob diferentes ataques FDI.

Tipo	Parâmetros	Erro médio [V]	y_{regime} [V]
Sem ataque	–	0,004	3,00
Aditivo	$a_0 = 0,2$ V	0,198	2,80
	$a_0 = 0,5$ V	0,495	2,50
	$a_0 = 1,0$ V	0,990	2,00
Multiplicativo	$\alpha = 0,8$	0,727	3,75
	$\alpha = 1,2$	0,496	2,50
	$\alpha = 1,5$	0,993	2,00
Senoidal	$A_a = 0,3$ V, $f_a = 2$ Hz	0,188	$3,0 \pm 0,30$

Para evitar detecção, o atacante pode projetar o sinal de ataque de forma a respeitar limites estatísticos. O conhecimento do modelo do sistema, obtido na fase de identificação, permite ao atacante calcular $\hat{y}(k)$ e projetar $a(k)$ de forma a permanecer abaixo do limiar. Os ataques senoidais de pequena amplitude são particularmente difíceis de detectar, pois podem ser confundidos com ruído de medição ou variações normais do processo.

B.2 Ataque de *Replay*

O ataque de *replay* consiste na gravação de uma sequência de medições legítimas $\{y(k), y(k+1), \dots, y(k+L-1)\}$ durante um período de operação normal do sistema, seguida da reprodução cíclica desses dados para mascarar atividades maliciosas. O algoritmo implementado no *software man-in-the-middle* opera em duas fases distintas:

B.2.1 Fase de Gravação

Durante L instantes de amostragem, o atacante armazena as medições legítimas em um *buffer* circular:

$$\mathcal{B} = \{y(k_0), y(k_0 + 1), \dots, y(k_0 + L - 1)\}, \quad (\text{B.6})$$

em que L representa o tamanho da janela de gravação e k_0 é o instante de início da gravação, escolhido conforme descrito a seguir.

B.2.2 Seleção do Ponto de Corte via Análise Espectral

A eficácia do ataque de *replay* depende da qualidade do segmento gravado. Uma reprodução cíclica que apresente descontinuidades ou variações significativas pode ser detectada por sistemas de monitoramento. Para auxiliar na seleção do ponto de corte, foram realizadas análises espectrais do sinal de saída. O procedimento de análise consiste nas seguintes etapas:

1. Análise de potência em janela deslizante: Calcula-se a potência instantânea do sinal utilizando uma janela deslizante de duração T_w :

$$P(k) = \frac{1}{N_w} \sum_{i=0}^{N_w-1} y^2(k+i), \quad (\text{B.7})$$

em que $N_w = T_w/T_s$ é o número de amostras na janela. Regiões de baixa variação de potência indicam operação em regime permanente.

2. Transformada de Fourier de Curta Duração (STFT): O espectrograma do sinal é calculado para visualizar o conteúdo frequencial ao longo do tempo:

$$S(f, \tau) = \left| \int_{-\infty}^{\infty} y(t) \cdot w(t - \tau) \cdot e^{-j2\pi ft} dt \right|^2, \quad (\text{B.8})$$

em que $w(t)$ é uma função janela. A Figura 45 apresenta o espectrograma do sinal de saída.

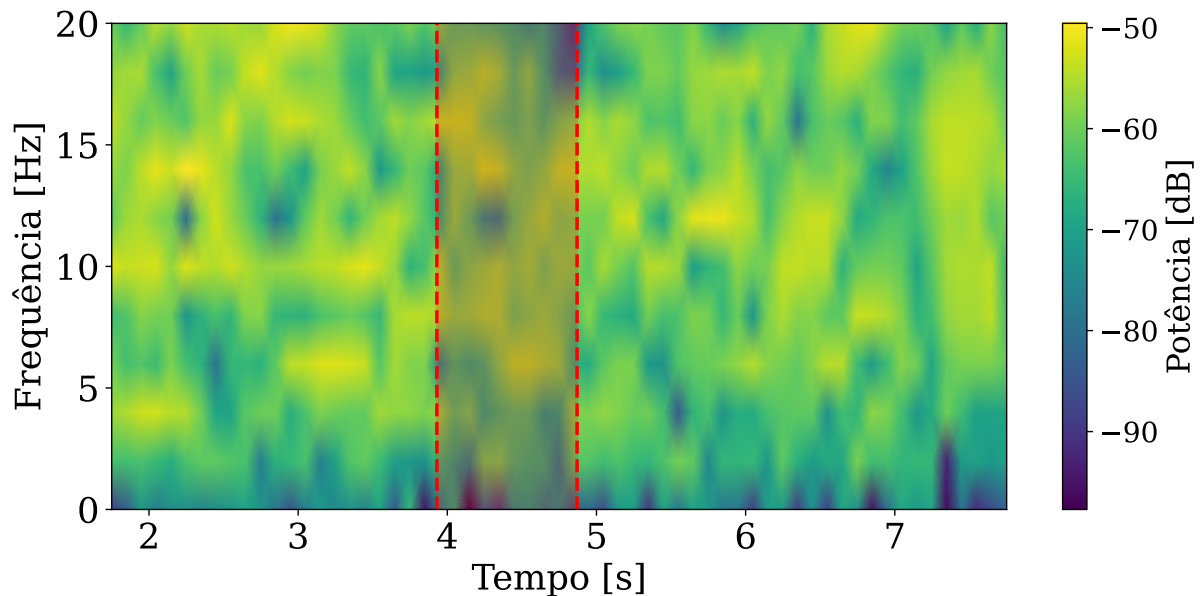


Figura 45 – Espectrograma do sinal de saída durante operação normal. A região delimitada pelas linhas tracejadas vermelhas indica a janela selecionada para gravação.

3. Identificação da frequência dominante: A partir da transformada de *Fourier* rápida (FFT) do sinal em regime permanente, identifica-se a frequência dominante f_d e

seu período correspondente $T_d = 1/f_d$. O tamanho do *buffer* é escolhido como múltiplo inteiro de T_d para facilitar a continuidade na reprodução cíclica:

$$L = \left\lceil \frac{T_{min}}{T_d} \right\rceil \cdot \frac{T_d}{T_s}, \quad (\text{B.9})$$

em que T_{min} é a duração mínima desejada para o *buffer*. Para o sistema em estudo, a análise revelou uma frequência dominante de $f_d = 12,77$ Hz, correspondendo a um período de $T_d = 78,3$ ms. O *buffer* foi dimensionado com $L = 94$ amostras (940 ms).

4. Seleção do ponto de início: Conforme ilustrado na Figura 46, o sinal de saída $y(t)$ apresenta variação muito pequena durante todo o regime permanente, mantendo-se próximo ao *setpoint* de 3 V. A potência instantânea $P(t)$ oscila em torno de 9 V^2 com amplitude de apenas $0,04 \text{ V}^2$.

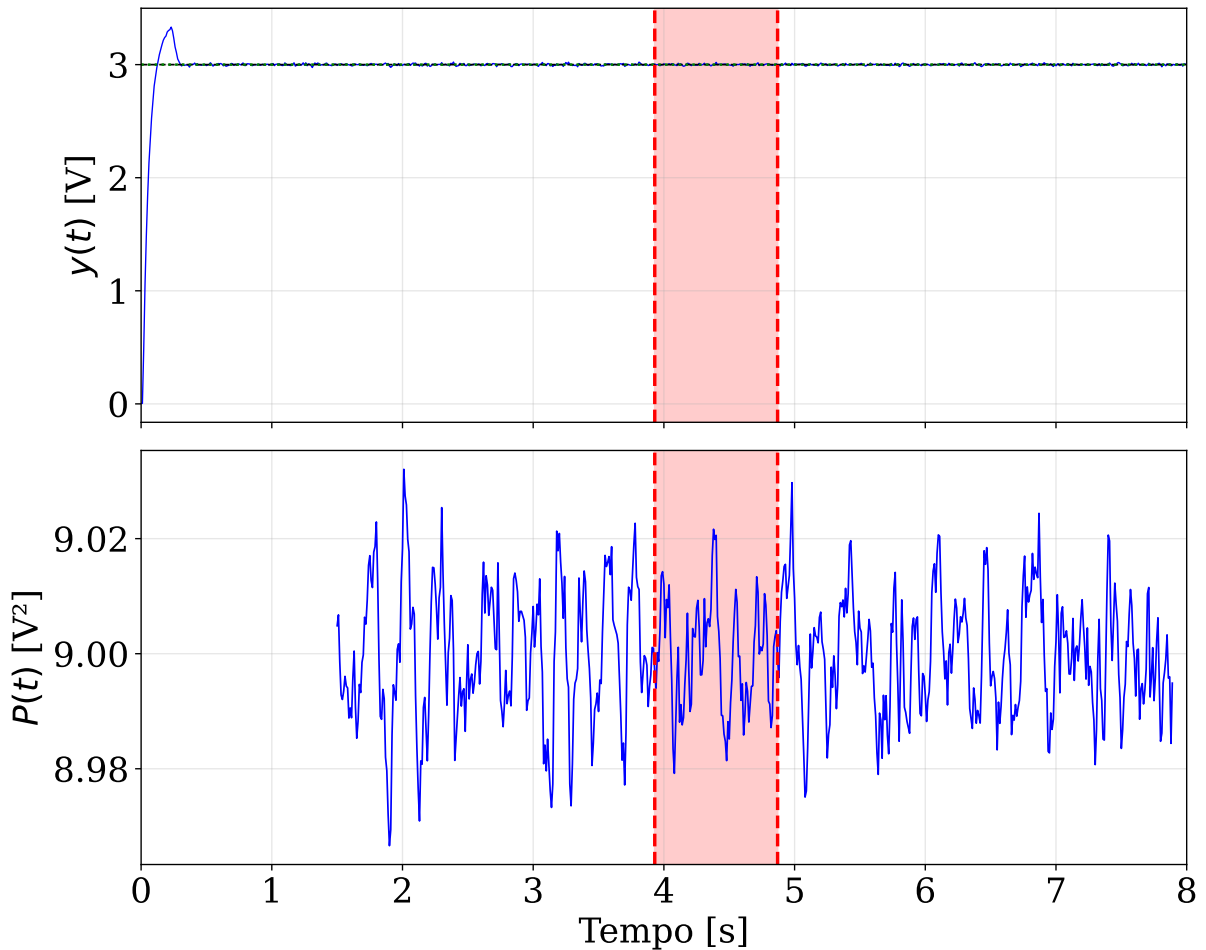


Figura 46 – Análise do sinal para seleção da janela de gravação: o painel superior mostra o sinal de saída $y(t)$ com indicação do *setpoint* e da média em regime permanente; e o painel inferior mostra a potência instantânea $P(t)$ calculada em janela deslizante de 100 ms.

Devido a essa característica do sistema em malha fechada com controlador LQR, a variância do sinal é uniformemente baixa em toda a região após o transitório. Consequentemente, qualquer janela selecionada após $t \approx 1,5$ s apresentaria características estatísticas

semelhantes. O ponto de início k_0 foi então escolhido de forma arbitrária dentro do regime permanente, em $t_0 = 3,93$ s, garantindo apenas que:

1. O transitório inicial já tenha passado;
2. O sistema esteja operando em regime estacionário;
3. Haja tempo suficiente após a janela de gravação para demonstrar a fase de reprodução.

Essa seleção arbitrária é justificada pela natureza do sistema, em um cenário com maior variabilidade no sinal de saída (como sistemas com perturbações periódicas ou referências variantes no tempo), a análise espectral seria essencial para identificar a janela ótima de gravação.

B.2.3 Fase de Reprodução

A partir do instante $k \geq k_0 + L$, o atacante substitui as medições reais pelas gravadas:

$$y_a(k) = \mathcal{B}[(k - k_0 - L) \bmod L] = y(k_0 + ((k - k_0 - L) \bmod L)), \quad (\text{B.10})$$

em que $y_a(k)$ representa a medição adulterada enviada ao controlador. Essa reprodução mantém o controlador “cego” para o estado real da planta.

B.2.4 Injeção de Sinal Malicioso na Ação de Controle

Simultaneamente à reprodução das medições, o atacante pode executar ações maliciosas na planta. Considerando que o canal de atuação não possui monitoramento independente (ausência de sensor medindo a ação de controle), o atacante injeta um sinal periódico $a(k)$ ao controle legítimo:

$$u_{real}(k) = u_{leg}(k) + a(k), \quad (\text{B.11})$$

em que $u_{leg}(k)$ é o comando calculado pelo controlador e $a(k)$ é um sinal de ataque. Nesse trabalho, utiliza-se uma onda quadrada de amplitude A_a e período T_a :

$$a(k) = A_a \cdot \text{sign} \left(\sin \left(\frac{2\pi k T_s}{T_a} \right) \right). \quad (\text{B.12})$$

Esse sinal periódico de alta frequência causa estresse mecânico no atuador, acelerando seu desgaste sem alterar o valor médio da ação de controle. A Figura 47 apresenta a comparação entre a saída vista pelo controlador (constante devido ao *replay*) e a saída real da planta (oscilante devido ao ataque).

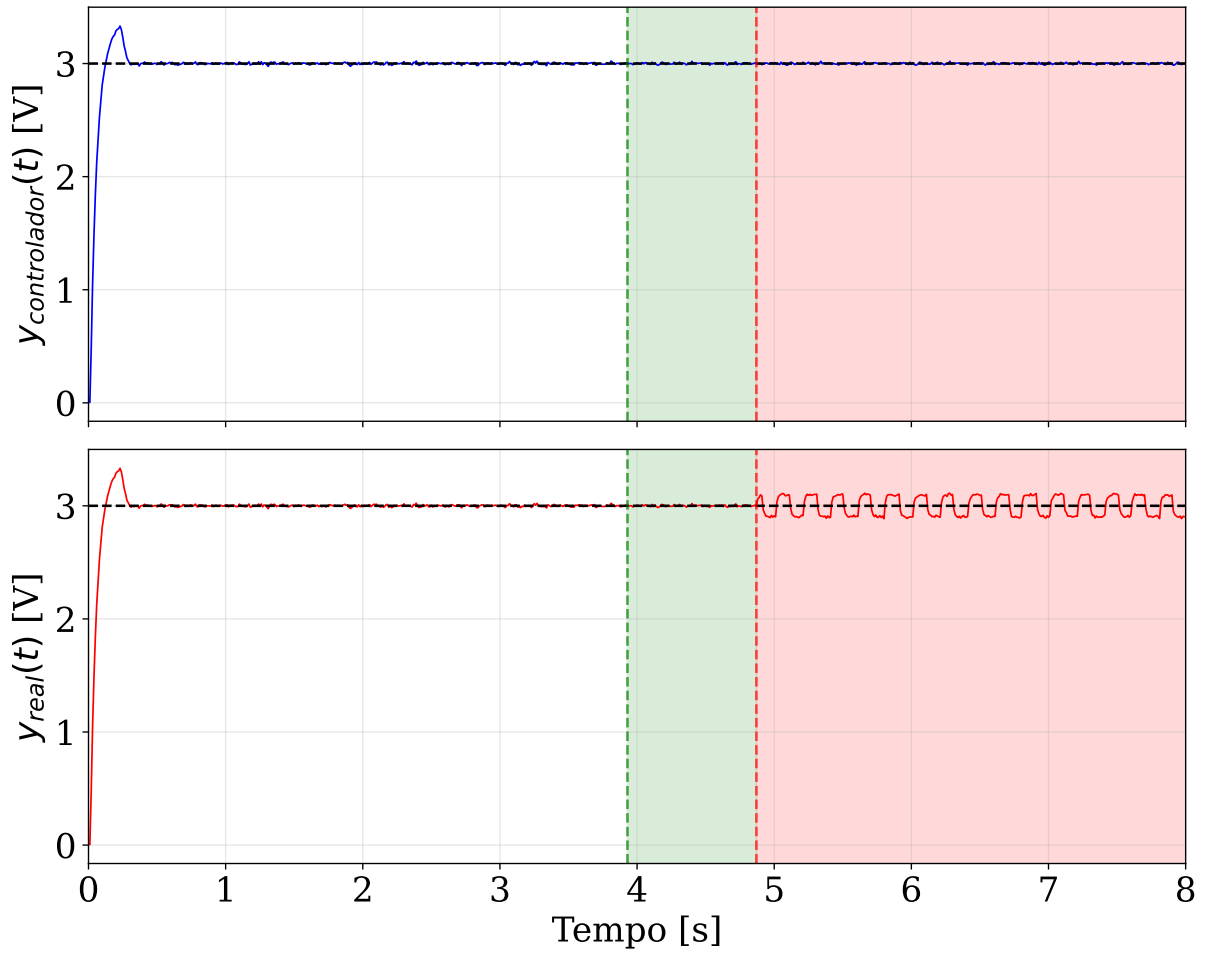


Figura 47 – Efeito do ataque de *replay* na saída do sistema: o painel superior mostra a saída percebida pelo controlador $y_{controlador}(t)$; e o painel inferior mostra a saída real da planta $y_{real}(t)$. A região verde indica a fase de gravação e a região vermelha indica a fase de reprodução.

A Figura 48 mostra a ação de controle calculada pelo controlador e a ação realmente aplicada à planta. Durante a fase de reprodução, o controlador tenta corrigir o erro percebido (baseado nas medições reproduzidas), enquanto a planta recebe o controle adulterado com o sinal de ataque sobreposto.

B.2.5 Impacto no Desgaste do Atuador

O objetivo secundário do ataque é causar degradação física do atuador por meio de ciclos de estresse mecânico. Durante a simulação do ataque, foram obtidos os seguintes resultados para a fase de reprodução:

- Variação Total do controle legítimo: $TV_{leg} = 46,30$
- Variação Total do controle real: $TV_{real} = 107,77$
- Incremento percentual: $\Delta TV = 132,8\%$

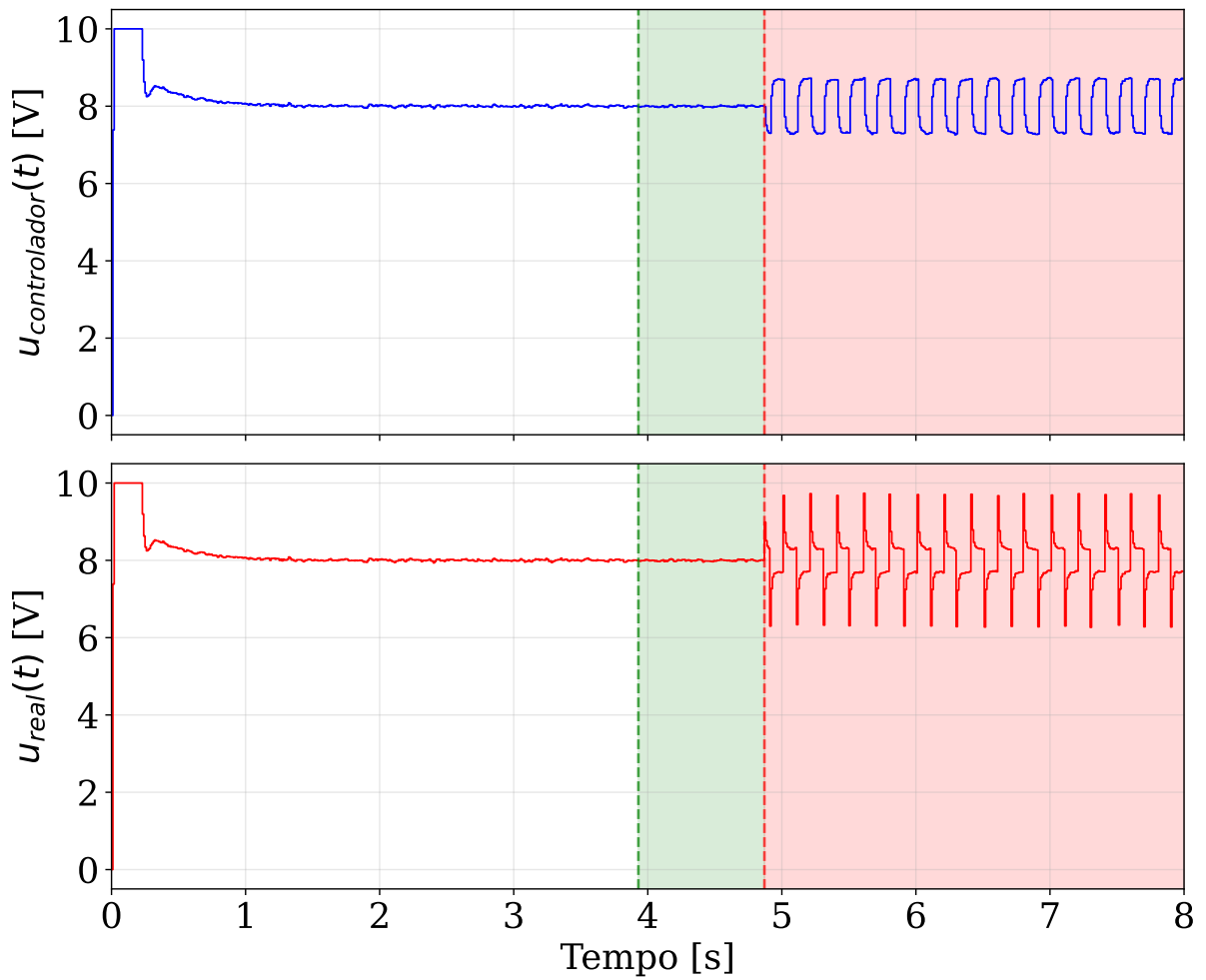


Figura 48 – Ação de controle durante o ataque de *replay*: o painel superior mostra o controle calculado pelo controlador $u_{\text{controlador}}(t)$; e o painel inferior mostra o controle realmente aplicado à planta $u_{\text{real}}(t)$.

Esse aumento na variação total indica que o atuador está sendo submetido a mais que o dobro das transições de estado que ocorreriam em operação normal, acelerando seu desgaste e reduzindo sua vida útil.

B.3 Ataque de Atraso

O ataque de atraso consiste na introdução de um retardo artificial τ_a no canal de comunicação, atrasando a chegada das medições ao controlador ou dos comandos à planta. Esse tipo de ataque pode desestabilizar sistemas de controle cuja estabilidade depende do sincronismo temporal.

B.3.1 Modelagem do Sistema com Atraso

Para avaliar a vulnerabilidade do sistema identificado a esse ataque, analisa-se o impacto do atraso na estabilidade em malha fechada. O sistema em malha fechada com

controlador LQR com ação integral e atraso de d períodos de amostragem na realimentação pode ser representado pelo sistema aumentado. O vetor de estados aumentado inclui os estados originais, os estados atrasados e o integrador:

$$\tilde{\mathbf{x}}(k) = \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{x}(k-1) \\ \vdots \\ \mathbf{x}(k-d) \\ x_i(k) \end{bmatrix}, \quad (\text{B.13})$$

em que $x_i(k)$ é o estado do integrador. A dinâmica do sistema aumentado é dada por:

$$\tilde{\mathbf{x}}(k+1) = \tilde{\mathbf{A}}\tilde{\mathbf{x}}(k), \quad (\text{B.14})$$

com a matriz $\tilde{\mathbf{A}}$ construída de forma que a lei de controle $u(k) = -\mathbf{K}_x\mathbf{x}(k-d) - K_i x_i(k)$ utilize o estado atrasado.

B.3.2 Influência do Ganho do Controlador na Robustez ao Atraso

A robustez do sistema ao ataque de atraso depende diretamente dos ganhos do controlador $\mathbf{K} = [\mathbf{K}_x \ K_i]$. No projeto LQR, o parâmetro R da função custo determina a “agressividade” do controle, valores menores de R resultam em ganhos maiores, proporcionando resposta mais rápida porém menor robustez a atrasos.

A Tabela 18 apresenta o atraso crítico d_{crit} (menor valor de d que torna o sistema instável) para diferentes valores de R , mantendo $\mathbf{Q} = \mathbf{I}_2$ e $Q_i = 100$.

Tabela 18 – Influência do ganho do controlador na robustez ao ataque de atraso.

R	$\mathbf{K}_x$	K_i	d_{crit}	τ_{crit} [ms]
1	[15,08; 28,35]	-5,94	1	10
10	[14,81; 21,54]	-2,46	2	20
50	[14,31; 18,08]	-1,23	4	40
100	[13,96; 16,81]	-0,90	6	60

Observa-se que controladores mais “agressivos” (menor R , maior ganho) são mais vulneráveis ao ataque de atraso. Esse comportamento é ilustrado na Figura 49, que mostra a evolução do máximo módulo dos autovalores $|\lambda|_{max}$ em função do atraso d para diferentes valores de R . O limiar de estabilidade $|\lambda|_{max} = 1$ é ultrapassado mais rapidamente para controladores com ganhos elevados.

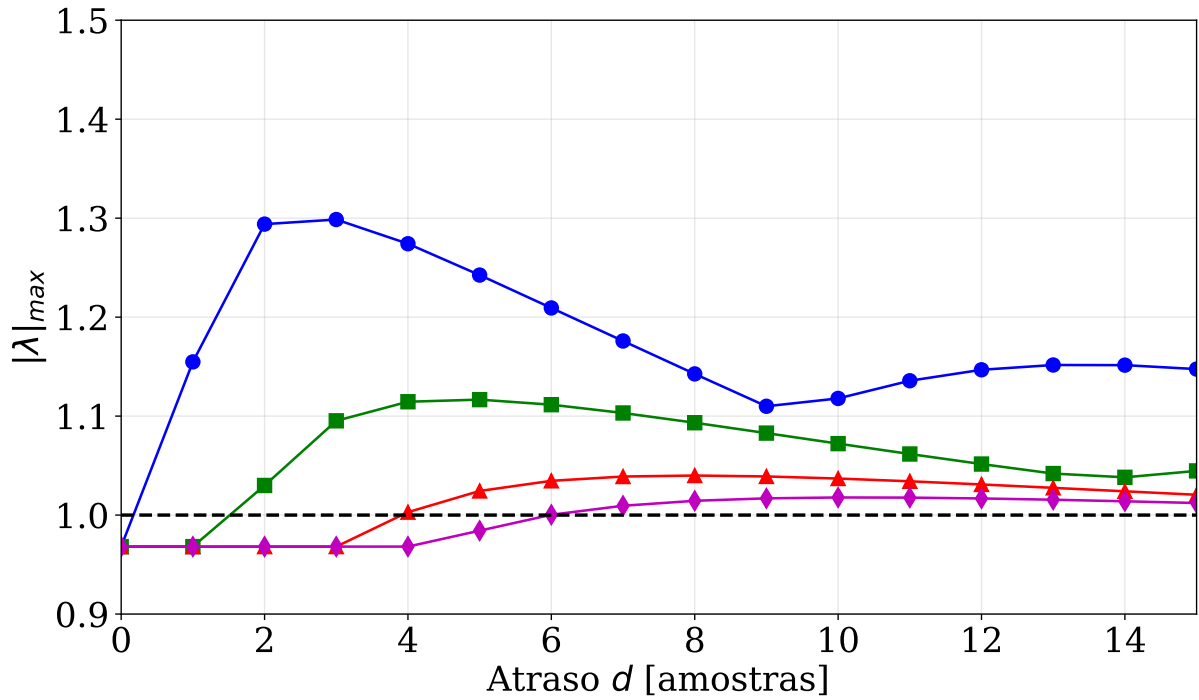


Figura 49 – Máximo módulo dos autovalores $|\lambda|_{max}$ do sistema aumentado em função do atraso d para diferentes valores do parâmetro R do LQR. Azul com círculos: $R = 1$; verde com quadrados: $R = 10$; vermelho com triângulos: $R = 50$; magenta com losangos: $R = 100$. A linha tracejada preta indica o limiar de estabilidade ($|\lambda|_{max} = 1$).

B.3.3 Análise para o Controlador Projetado

Para o controlador utilizado nesse trabalho ($R = 10$, $\mathbf{Q} = \mathbf{I}_2$, $Q_i = 100$), os ganhos obtidos são:

$$\mathbf{K}_x = \begin{bmatrix} 14,81 & 21,54 \end{bmatrix}, \quad K_i = -2,46. \quad (\text{B.15})$$

Os autovalores do sistema em malha fechada sem atraso são $\lambda = \{0,968; 0,660 \pm 0,216j\}$, todos dentro do círculo unitário. A Tabela 19 apresenta a análise de estabilidade para valores crescentes de atraso.

Tabela 19 – Análise de estabilidade do sistema com atraso ($R = 10$).

d [amostras]	Atraso [ms]	$ \lambda _{max}$
0	0	0,968
1	10	0,968
2	20	1,030
3	30	1,095
4	40	1,115
5	50	1,117

O sistema torna-se instável para $d \geq 2$ amostras, correspondendo a um atraso crítico de $\tau_{crit} = 20$ ms. A Figura 50 apresenta a mudança dos autovalores no plano

complexo conforme o atraso aumenta, evidenciando a saída do círculo unitário para $d \geq 2$.

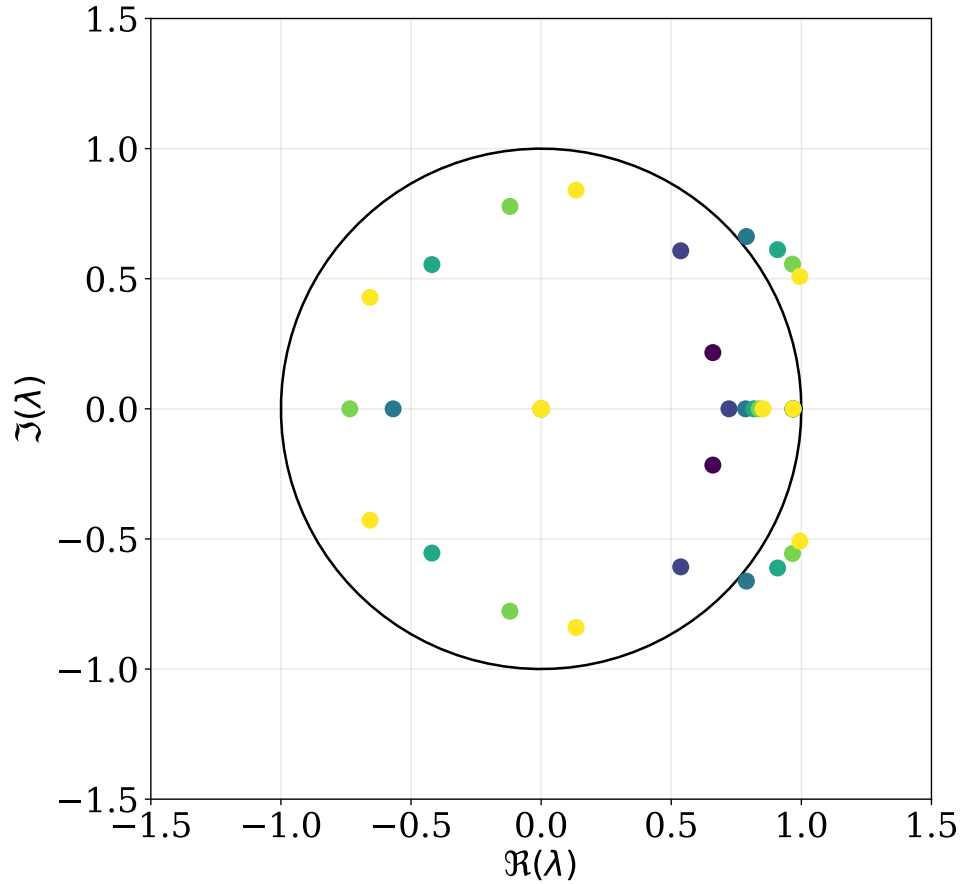


Figura 50 – Localização dos autovalores do sistema aumentado no plano complexo para diferentes valores de atraso d . Roxo escuro: $d = 0$; azul escuro: $d = 1$; ciano: $d = 2$; verde: $d = 3$; verde claro: $d = 4$; amarelo: $d = 5$. A circunferência preta representa o círculo unitário (limiar de estabilidade).

B.3.4 Implementação do Ataque

O atacante implementa o ataque armazenando as medições em uma fila de tamanho d :

$$y_a(k) = y(k - d), \quad (\text{B.16})$$

em que $y_a(k)$ é a medição atrasada enviada ao controlador. Na prática, o *software* de ataque mantém um *buffer* circular que recebe as medições atuais e libera as medições antigas com o atraso desejado.

A Figura 51 apresenta a resposta temporal do sistema sob ataque de atraso iniciado em $t = 1,5$ s. Observa-se que:

- Para $d = 0$ (sem ataque) e $d = 1$ (10 ms), o sistema permanece estável com a saída convergindo para o *setpoint*;

- Para $d = 2$ (20 ms) e $d = 3$ (30 ms), o sistema torna-se instável, apresentando oscilações crescentes tanto na saída quanto na ação de controle.

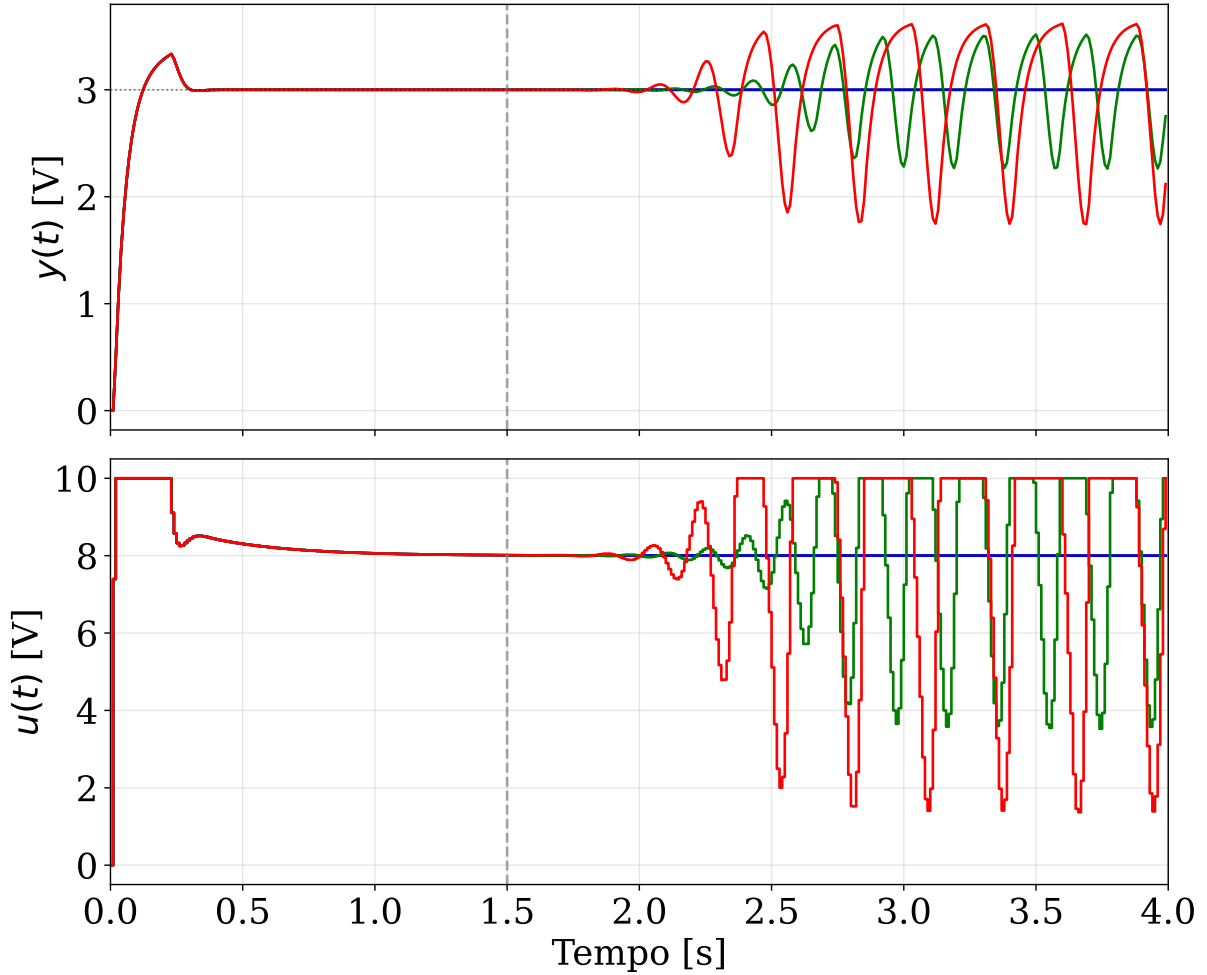


Figura 51 – Resposta temporal do sistema sob ataque de atraso iniciado em $t = 1,5$ s (linha vertical tracejada cinza): o painel superior mostra a saída $y(t)$; e o painel inferior mostra a ação de controle $u(t)$. Preto: sem ataque ($d = 0$); azul: $d = 1$ amostra (10 ms); verde: $d = 2$ amostras (20 ms); vermelho: $d = 3$ amostras (30 ms).

Observa-se a instabilidade na forma de oscilações crescentes que forçam a saturação da ação de controle (0 V e 10 V). Esse comportamento oscilatório é característico de sistemas com atraso na malha de realimentação, onde o controlador reage a informações defasadas, amplificando os erros ao invés de corrigi-los.