Uso de Controle Chaveado para Mitigação de Ataque Ativo de Identificação de Sistemas com Malha Fechada

Alan Oliveira de Sá[1,2], **Luiz F. Rust da Costa Carmo**[1,3], **Raphael C. S. Machado**[3,4]

[1]Programa de Pós-Graduação em Informática - Instituto Tércio Pacitti / IM,
Universidade Federal do Rio de Janeiro, 21.941-901, RJ – Brasil

[2]Centro de Instrução Almirante Wandenkolk – Marinha do Brasil,
Ilha das Enxadas, Baía de Guanabara – Rio de Janeiro – RJ – Brasil

[3]Instituto Nacional de Metrologia, Qualidade e Tecnologia (Inmetro)
Av. Nossa Senhora das Graças, 50, Xerém, Duque de Caxias, 25.250-020, RJ – Brasil

[4]Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ)
Av. Maracanã, 229, Maracanã, 20.271-110 - Rio de Janeiro – RJ – Brasil

alan.oliveira.sa@gmail.com, {lfrust,rcmachado}@inmetro.gov.br

Abstract. *The literature regarding to cyber-physical attacks in Networked Control Systems (NCS) indicates that covert and accurate attacks must be planned based on an accurate knowledge about the model of the attacked system. In this sense, the literature on NCS recognizes the Active System Identification attack as a tool to provide the attacker with the required system models. However, there is still a lack of discussion about countermeasures for this specific attack. In this sense, this work proposes the use of a randomly switching controller as a countermeasure for the Active System Identification attack. The simulation results indicate that this countermeasure is capable to mitigate the mentioned attack at the same time that it performs a satisfactory plant control.*

Resumo. *A literatura referente aos ataques físico-cibernéticos em Sistemas de Controle em Rede indica que ataques furtivos e acurados devem ser planejados com base em um conhecimento igualmente acurado sobre o modelo do sistema atacado. Neste contexto, a literatura sobre Sistemas de Controle em Rede reconhece o ataque Ativo de Identificação de Sistemas como uma ferramenta capaz de prover tais modelos ao atacante. No entanto, há uma carência de discussão sobre contramedidas para este ataque específico. Nesse sentido, este trabalho propõe o uso de controladores aleatoriamente chaveados como uma contramedida para o ataque Ativo de Identificação de Sistemas. Os resultados de simulação indicam que esta contramedida é capaz de mitigar o referido ataque e executar, simultaneamente, um controle satisfatório da planta.*

1. Introdução

Um Sistema de Controle em Rede, ou *Networked Control System* (NCS), consiste em uma planta física controlada por um controlador digital – *i.e.* um sistema computacional – por meio de uma rede de comunicação que, deste modo, integra o ciberespaço ao domínio físico. Considerando o crescente uso de NCSs em plantas industriais e infraestruturas críticas, bem como as ameaças cibernéticas que podem afetar estes sistemas, estudos têm sido realizados para caracterizar vulnerabilidades e propor soluções de segurança para tais sistemas. Neste contexto, a literatura [Smith 2011, Amin et al. 2013, Teixeira et al. 2015,

Smith 2015, de Sá et al. 2017b] demonstra que uma série de ataques sofisticados a NCSs requerem um conhecimento prévio sobre o modelo do sistema atacado. Considerando esta necessidade, recentes trabalhos [de Sá et al. 2017b, de Sá et al. 2017a] apresentam um conjunto de ataques de Identificação de Sistemas que podem ser lançados contra NCSs para fornecer ao atacante os requeridos modelos e, portanto, subsidiar a elaboração de outros ataques.

A literatura [de Sá et al. 2017b, Smith 2011, Smith 2015, Teixeira et al. 2015] indica que, para ser eficaz, ataques furtivos e acurados devem contar com um conhecimento igualmente acurado sobre o modelo do sistema atacado. Em [de Sá et al. 2017b], por exemplo, a operação conjunta de um ataque de Identificação de Sistemas e um ataque de Injeção de Dados é utilizada para degradar, de forma fisicamente furtiva, o serviço desempenhado por uma planta. No referido trabalho, é demonstrado que a performance deste ataque furtivo de injeção de dados é diretamente afetado pela acurácia da informação obtida pelo ataque de Identificação de Sistemas.

Conforme mostrado na Figura 1, o conhecimento sobre o sistema é um dos requisitos para o lançamento de uma série de ataques dependentes de modelo, tais como ataques controlados de *Denial of Service* (DoS-Controlled) [de Sá et al. 2017b] e ataques controlados de *Service Degradation* (SD-Controlled) [de Sá et al. 2017b]. Nesta figura, é possível ver que o Ataque Ativo de Identificação de Sistemas, ou *Active System Identification*, proposto em [de Sá et al. 2017a] constitui um caminho para a implementação dos referidos ataques, uma vez que ele é capaz de proporcionar ao atacante o conhecimento necessário sobre o sistema. Embora os autores de [de Sá et al. 2017a] encorajem o desenvolvimento de contramedidas para o Ataque Ativo de Identificação de Sistemas, há carência de discussão sobre contramedidas para este ataque específico. Neste sentido, este trabalho tem por objetivo discutir e propor uma contramedida para o referido ataque.

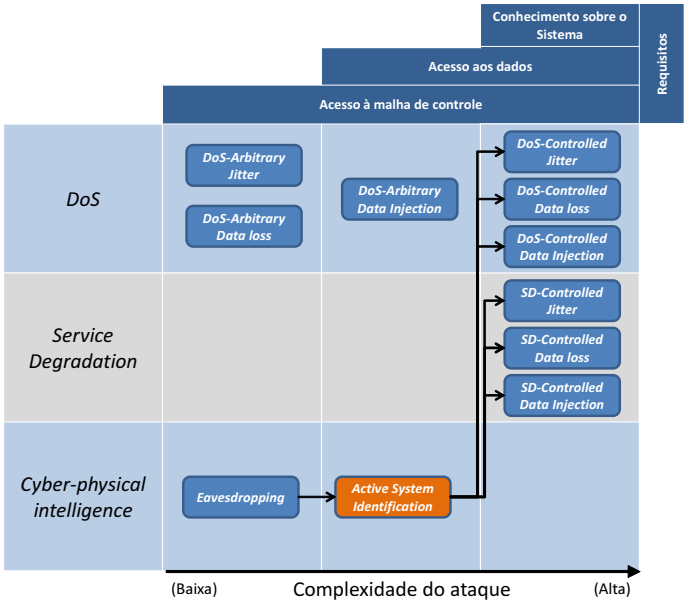


Figura 1. Categorias e requisitos de ataques em NCSs.

A contramedida mais evidente para prevenir o sucesso de um ataque de Identificação de Sistemas em um NCS é prevenir o acesso não autorizado à malha de controle utilizando, por exemplo, segmentação de rede, *demilitarized zones* (DMZ), políticas de *firewall* e implementando arquiteturas de rede específicas, tal como estabe-

lecido em [Stouffer et al. 2015]. Uma contramedida complementar – para o caso de o atacante conseguir acessar a malha de controle – é prevenir o acesso aos dados que fluem no NCS utilizando, por exemplo, algoritmos de criptografia de chaves simétricas, algoritmos de *hash* e uma estratégia de *timestamp* para formar um mecanismo seguro de comunicação entre o controlador e a planta, tal como proposto em [Pang and Liu 2012]. Entretanto, quando as referidas contramedidas falham e o atacante obtém acesso aos dados que fluem no NCS, a alternativa para prevenir que o atacante obtenha o modelo do sistema é dificultar a análise do dado capturado – *i.e.* tornar o algoritmo de Identificação de Sistemas impreciso/ineficaz.

Uma possível estratégia para causar dificuldades ao algoritmo de Identificação de Sistemas é ter no NCS funções de controle específicas que sejam, ao mesmo tempo, mais difíceis de serem identificadas e capazes de controlar a planta. Considerando esta estratégia, é proposto neste trabalho o uso de controladores aleatoriamente chaveados como uma contramedida factível para o Ataque Ativo de Identificação de Sistemas, proposto em [de Sá et al. 2017a].

O restante deste artigo está organizado da seguinte forma. Na Seção 2, são apresentados trabalhos relacionados à esta pesquisa. Na Seção 3, o ataque Ativo de Identificação de Sistemas proposto em [de Sá et al. 2017a] é brevemente descrito. Na Seção 4, o controlador chaveado é apresentado e discutido como uma contramedida para o Ataque Ativo de Identificação de Sistemas. A Seção 5 apresenta resultados de simulações, onde o desempenho do controlador chaveado é analisado do ponto de vista da contramedida e do controle. Finalmente, na Seção 6, são apresentadas as conclusões e possíveis trabalhos futuros.

2. Trabalhos Relacionados

Nesta seção é apresentada uma revisão sobre trabalhos relacionados aos ataques físico-cibernéticos em NCSs. O foco está em estudos recentes que englobam ataques furtivos e dependentes de modelo, bem como ataques de Identificação de Sistemas.

Em [Teixeira et al. 2015], os autores analisam diversos ataques em NCSs e estabelecem que a implementação de ataques furtivos requer um elevado conhecimento sobre o modelo do sistema atacado. Em [Smith 2011, Amin et al. 2013, Smith 2015], são propostos e analisados ataques furtivos que corroboram o requisito estabelecido por [Teixeira et al. 2015]. Em [Smith 2011, Smith 2015], o atacante, atuando como um *Man-in-the-Middle* (MitM), injeta dados falsos nos sinais de controle do NCS para assumir o controle da planta. O atacante, então, utiliza o modelo da planta atacada para calcular os dados que são injetados no sinal de realimentação do NCS, a fim de tornar o ataque furtivo. A furtividade do ataque proposto em [Smith 2011] é analisada do ponto de vista dos sinais que chegam ao controlador e, conforme demonstrado em [Smith 2015], depende da diferença entre o modelo real da planta e o modelo conhecido pelo atacante. Em [Amin et al. 2013] o atacante, conhecendo o modelo do sistema atacado, injeta dados falsos no NCS para furtivamente roubar água do canal Gignac, localizado no sul da França.

Em [Amin et al. 2013, Smith 2011, Smith 2015, Teixeira et al. 2015], onde os ataques são planejados e implementados com base nos modelos dos sistemas atacados, não é descrito como estes modelos são obtidos pelo atacante. Considera-se apenas que os modelos são previamente conhecidos e, então, utilizados para subsidiar o planejamento destes ataques furtivos/dependentes de modelo.

Para preencher este hiato, em [de Sá et al. 2017a, de Sá et al. 2017b], os autores propõem dois novos tipos de ataque: o Ataque Passivo de Identificação de Sistemas, ou *Passive System Identification Attack* [de Sá et al. 2017b]; e o Ataque Ativo de Identificação de Sistemas, ou *Active System Identification Attack* [de Sá et al. 2017a]. Estes ataques, que pertencem à categoria de Ataques de Inteligência Físico-cibernética, ou *Cyber-physical Intelligence Attacks* [de Sá et al. 2017b], visam estimar o modelo dos sistemas atacados. O Ataque Passivo de Identificação de Sistemas [de Sá et al. 2017b] não precisa injetar sinais no NCS para estimar seu modelo. Entretanto, o mesmo depende da ocorrência de eventos, que não são controlados pelo atacante, para produzir sinais que carreguem informações significativas para o algoritmo de identificação. Tal ataque estima a função de transferência de cada dispositivo do NCS – *i.e.* controlador e planta – capturando passivamente os sinais de controle e de realimentação do sistema. O Ataque Ativo de Identificação de Sistemas [de Sá et al. 2017a], por sua vez, constitui uma alternativa aos Ataques Passivos de Identificação de Sistemas em situações em que o atacante não pode aguardar muito tempo pela ocorrência de sinais que contenham informações significativas para o processo de identificação. Para isso, conforme descrito na Seção 3, o atacante estima a função de transferência de malha aberta do NCS injetando um sinal de ataque e capturando a sua consequente resposta. Ambas as ações – *i.e.* injeção e captura de dados – são realizadas em um único ponto de interceptação do sistema. Neste trabalho é proposta uma contramedida para mitigar o Ataque Ativo de Identificação de Sistemas, mesmo que o atacante obtenha acesso aos dados transmitidos no NCS.

3. O Ataque Ativo de Identificação de Sistemas

Nesta Seção, o ataque Ativo de Identificação de Sistemas proposto em [de Sá et al. 2017a] é brevemente descrito, a fim de prover as informações basilares e necessárias à compreensão da contramedida proposta no presente trabalho. O referido ataque visa estimar os coeficientes da função de transferência de malha aberta $G(z) = C(z)P(z)$ de um NCS, apresentado na Figura 2, onde $C(z)$ é a função de controle executada pelo controlador e $P(z)$ é o modelo da planta. Para tal, o ataque é realizado em três etapas:

- ETAPA-I: Primeiramente, agindo como um *Man-in-the-Middle* (MitM), o atacante injeta um sinal de ataque $a(k)$ no NCS, conforme apresentado na Figura 2.
- ETAPA-II: Após injetar $a(k)$, o atacante captura a saída $y(k)$ da planta, durante um período de monitoração T, a fim de obter a resposta $y_a(k)$ causada por $a(k)$.
- ETAPA-III: Conhecendo o sinal de ataque $a(k)$ e a consequente resposta $y_a(k)$ do NCS, o atacante estima a função de transferência de malha aberta do sistema $G(z)$ aplicando $a(k)$ em um modelo estimado $G_e(z)$, o qual é ajustado até que a saída estimada $\hat{y}_a(k)$ coincida com $y_a(k)$. Em [de Sá et al. 2017a], este ajuste é realizado por algoritmos de otimização bio-inspirados.

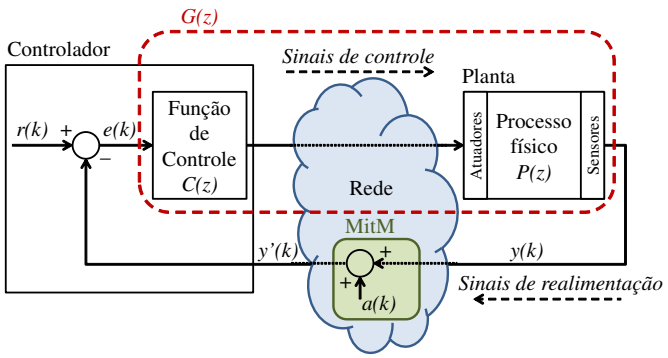


Figura 2. Ataque Ativo de Identificação de Sistema em NCS. [de Sá et al. 2017a]

Note que $y_a(k)$, obtido na ETAPA-II, é apenas uma parcela de $y(k)$. A resposta completa do sistema é $y(k) = y_r(k) + y_a(k)$, onde $y_r(k)$ é a resposta do sistema causada por $r(k)$. Considerando que o sistema é estável, a parcela $y_r(k)$ converge e estabiliza em um valor constante q após um determinado número de amostras k_s. Assim, na ETAPA-II, para obter $y_a(k)$, $\forall k > k_s$, o atacante deve calcular $y_a(k) = y(k) - q$. Isto elimina a parcela de $y(k)$ causada por $r(k)$, tornando o problema de identificação dependente de apenas o sinal de ataque $A(z) = \mathcal{Z}[a(k)]$, conforme descrito por (1):

$$y_a(k) = y(k) - q = -\mathcal{Z}^{-1}\left[\frac{G(z)}{1+G(z)}A(z)\right], \forall k > k_s. \quad (1)$$

onde $\mathcal{Z}$ representa a operação da transformada Z. O valor de q é obtido pelo atacante através da captura de $y(k)$ após a estabilização do NCS, antes da injeção de $a(k)$.

Na ETAPA-III, o atacante estima $G(z)$ aplicando $a(k)$ em um modelo estimado, definido por (2):

$$\hat{y}_a(k) = -\mathcal{Z}^{-1}\left[\frac{G_e(z)}{1+G_e(z)}\right] * a(k), \quad (2)$$

onde $G_e(z)$ é a estimativa de $G(z)$ e $\hat{y}_a(k)$ é a saída do sistema estimado. A função de transferência genérica $G_e(z)$, é definida como (3):

$$G_e(z) = \frac{\alpha_n z^n + \alpha_{n-1}z^{n-1} + ... + \alpha_1 z^1 + \alpha_0}{z^m + \beta_{m-1}z^{m-1} + ... + \beta_1 z^1 + \beta_0}, \quad (3)$$

onde $[\alpha_n, \alpha_{n-1}, ...\alpha_1, \alpha_0]$ e $[\beta_{m-1}, \beta_{m-2}, ...\beta_1, \beta_0]$ são os coeficientes do numerador e do denominador, respectivamente, os quais o Ataque Ativo de Identificação de Sistemas visa descobrir. A ordem do numerador e do denominador são expressas por n e m, respectivamente. Deste modo, para encontrar $G(z)$, os coeficientes de $G_e(z)$ são ajustados até que a saída estimada $\hat{y}_a(k)$ convirja para a resposta $y_a(k)$, conhecida.

Em [de Sá et al. 2017a], o *Backtracking Search Optimization Algorithm* (BSA) [Civicioglu 2013] e o *Particle Swarm Optimization* (PSO) [Kennedy 1995] são utilizados para ajustar iterativamente os parâmetros de $G_e(z)$, minimizando uma função de aptidão específica, até que $G_e(z)$ convirja para $G(z)$. Os coeficientes de $G_e(z)$ são as coordenadas $x_j = [\alpha_{n,j}, \alpha_{n-1,j}, ...\alpha_{1,j}, \alpha_{0,j}, \beta_{m-1,j}, \beta_{m-2,j}, ...\beta_{1,j}, \beta_{0,j}]$ de um indivíduo j do BSA/PSO. A aptidão f_j de cada indivíduo j do BSA/PSO é calculada de acordo com (4):

$$f_j = \frac{\sum_{k=0}^{N}(y_a(k) - \hat{y}_{aj}(k))^2}{N}, \quad (4)$$

onde N é o número de amostras que existem durante o período de monitoração T da ETAPA-II; e $\hat{y}_{aj}(k)$ é a resposta do modelo estimado (2) causada por $a(k)$, quando os coeficientes de $G_e(z)$ são x_j. Se nenhuma outra entrada – perturbação ou ruído – afetar o NCS durante T, então $\min f_j = 0$ quando $[\alpha_{n,j}, \alpha_{n-1,j}, ...\alpha_{1,j}, \alpha_{0,j}, \beta_{m-1,j}, \beta_{m-2,j}, ...\beta_{1,j}, \beta_{0,j}] = [\alpha_n, \alpha_{n-1}, ...\alpha_1, \alpha_0, \beta_{m-1}, \beta_{m-2}, ...\beta_1, \beta_0]$, *i.e.* quando $G_e(z)$ converge para $G(z)$.

4. Controladores Chaveados: uma Contramedida

Conforme discutido na Seção 1, uma possível forma de causar dificuldades ao algoritmo de identificação de sistemas é ter, no NCS, funções de transferências específicas que sejam mais difíceis de ser identificadas. Sendo assim, é necessário se debruçar sobre as

duas funções de transferência $C(z)$ e $P(z)$, apresentadas na Figura 2, para verificar o que pode ser feito para dificultar a identificação do NCS. No caso da planta, não é desejável ou mesmo viável modificar a sua função de transferência $P(z)$ apenas para torná-la mais difícil de identificar. Isto decorre do simples fato de que a função de transferência da planta é uma consequência da estrutura física da mesma. Portanto, modificar $P(z)$ significa modificar o processo físico que está sendo controlado, o que não é conveniente. Entretanto, é possível projetar uma função de controle de forma que esta atenda, simultaneamente, dois objetivos:

Objetivo I - Cumprir os requisitos de controle da planta considerando, em primeiro lugar, a estabilidade do sistema e, em segundo lugar, outros requisitos, tais como: tempo de acomodação, *overshoot*, etc.

Objetivo II - Dificultar o processo de identificação, fazendo com que o modelo obtido pelo atacante seja impreciso ou ambíguo, de tal forma que o atacante hesite em lançar contra o NCS ataques furtivos ou dependentes de modelo.

Note que, no caso do Ataque Ativo de Identificação de Sistemas proposto em [de Sá et al. 2017a], o atacante não identifica $C(z)$ e $P(z)$ separadamente, analisando suas respectivas entradas e saídas. O atacante intercepta a malha de controle em apenas um ponto e, a partir deste ponto de interceptação, estima a função de transferência de malha aberta do sistema $G(z) = C(z)P(z)$, conforme mostra a Figura 2. Assumindo que não é conveniente modificar $P(z)$, conforme anteriormente discutido, $C(z)$ deve ser projetado para dificultar a identificação da função de transferência de malha aberta $G(z) = C(z)P(z)$ do NCS. Neste sentido, considerando os dois objetivos apresentados, é proposto neste artigo o uso de controladores chaveados como uma contramedida para o Ataque Ativo de Identificação de Sistemas.

Um controlador chaveado consiste em um conjunto de funções de controle $C_i(z)$, $i \in \mathcal{I} = \{1, ..., N\}$, comutadas entre N estados por meio de uma regra de comutação S, para controlar uma planta $P(z)$, conforme apresentado na Figura 3. Quando as funções de controle $C_i(z)$ e a função de transferência da planta $P(z)$ são lineares, como no caso do presente artigo, o sistema é classificado como um sistema linear comutado, ou *switched linear system* (SLS). Por uma questão de simplicidade, mas sem prejuízo à generalidade, neste trabalho, o controlador chaveado é representado e discutido com apenas duas funções de controle $C_1(z)$ e $C_2(z)$ – *i.e.* $i \in \mathcal{I} = \{1, 2\}$.

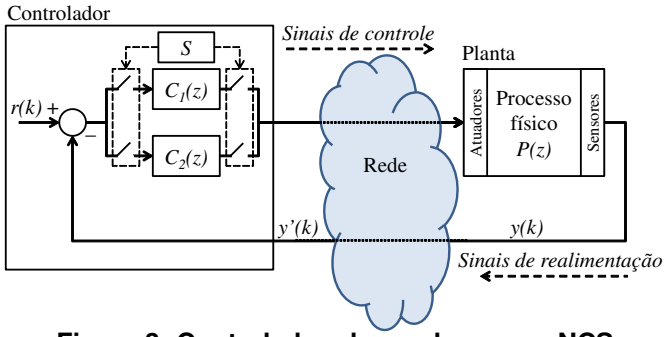


Figura 3. Controlador chaveado em um NCS.

Em geral, a regra de comutação S considera o comportamento da planta para chavear entre as funções de controle, tal como descrito em [Skafidas et al. 1999]. Entretanto, na solução proposta neste trabalho, a regra de comutação não observa o comportamento da planta para comandar os chaveamentos. Para tornar a identificação mais difícil, a regra de

comutação proposta ocorre segundo a cadeia de Markov da Figura 4, onde as funções de controle são chaveadas em intervalos aleatórios seguindo as probabilidades $p_{11}(l)$, $p_{12}(l)$. $p_{21}(l)$ e $p_{12}(l)$, onde l é o número de intervalos de amostragem ocorridos desde o último chaveamento. A razão para chavear as funções de controle em intervalos aleatórios é que, conforme [Wang 2013], se o tempo de cada chaveamento é conhecido, a identificação de um SLS se torna simples. No entanto, quando o tempo de cada chaveamento não é conhecido, a identificação de um SLS se torna uma tarefa não trivial. As probabilidades, $p_{12}(l)$ e $p_{21}(l)$ são obtidas da função de densidade de probabilidade (FDP) da Figura 5, onde a é o número mínimo de intervalos de amostragem que o sistema deve permanecer no mesmo estado e b é o número máximo de intervalos de amostragem que o sistema pode permanecer no mesmo estado. Note que $p_{11}(l) = 1 - p_{12}(l)$ e $p_{22}(l) = 1 - p_{21}(l)$.

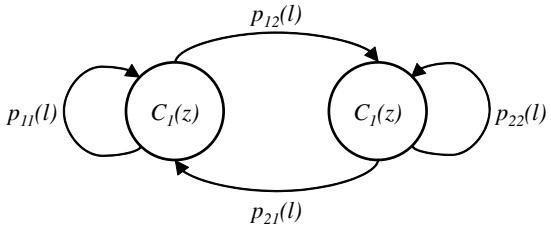


Figura 4. Regra de comutação pela cadeia de Markov.

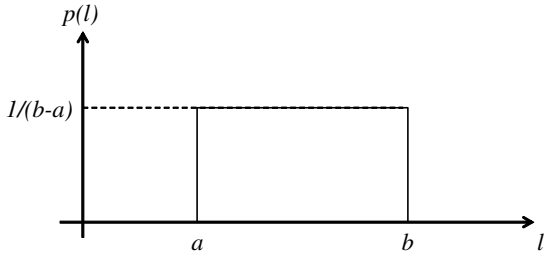


Figura 5. FDP de p_{12} e p_{21}.

Controladores com chaveamento estocástico, modelados por cadeias de Markov, já existem na literatura [Yang et al. 2011] com o objetivo de melhorar a estabilização e a resposta dinâmica do NCS em face de longos atrasos aleatórios na rede. No entanto, no presente trabalho – diferente de [Yang et al. 2011] –, as probabilidades de transição e os estados da cadeia são dimensionados para mitigar um ataque de identificação de sistemas mantendo a qualidade do controle.

Para ser estável sob comutações arbitrárias e sem restrições, o SLS deve atender a duas condições [Lin and Antsaklis 2009, Dasgupta et al. 2013]:

1. Todos os subsistemas – onde cada subsistema corresponde à função de transferência da planta $P(z)$ disposta em malha fechada com uma das funções de controle $C_i(z)$ – devem ser assintoticamente estáveis; e
2. Deve existir uma função de Lyapunov que seja comum para todos os subsistemas.

Assim, para tornar o SLS estável sob comutações arbitrárias e sem restrições, todas as funções de controle $C_i(z)$ devem ser projetadas de forma a atender às duas condições acima descritas.

Outra estratégia válida para alcançar a estabilidade em um SLS é restringir os eventos de chaveamento estabelecendo, por exemplo, um tempo mínimo entre dois chaveamentos consecutivos – *i.e.* um mínimo *dwell time*. Em um SLS, a instabilidade gerada ao chavear entre dois subsistemas estáveis é causada pela incapacidade de absorver o aumento de energia gerado pelas sucessivas comutações [Lin and Antsaklis 2009]. Intuitivamente, é razoável pensar que se o sistema permanecer nos subsistemas estáveis por um tempo suficiente – usando uma regra de comutação lenta – é possível prevenir o aumento de energia causado pelos chaveamentos. Com base neste conceito, em [Morse 1996], é provado que a estabilidade global de um SLS é garantida quando todos os seus subsistemas são estáveis e o *dwell time* é suficientemente grande. No entanto, é factível que o SLS ocasionalmente tenha um *dwell time* menor, desde que isso não ocorra com

frequência. Conforme apresentado em [Hespanha and Morse 1999], se todos os subsistemas forem exponencialmente estáveis, então o SLS será estável desde que o tempo médio de comutação seja suficientemente grande. Este conceito, referido como *average dwell-time*, é estendido em [Zhai et al. 2002] para SLSs discretos no tempo – que é o caso de NCSs com controle chaveado.

Neste trabalho, ao invés de projetar $C_1(z)$ e $C_2(z)$ para que o SLS seja estável sob comutações arbitrárias e sem restrições – *i.e.* atendendo às condições 1 e 2, previamente descritas – utiliza-se a estratégia de restringir as comutações do SLS. Neste sentido, primeiramente, $C_1(z)$ e $C_2(z)$ são dimensionadas de forma independente com base na análise do lugar das raízes, a fim de que cada subsistema seja estável. A estabilidade global do SLS é, então, alcançada ajustando os parâmetros a e b da FDP apresentada na Figura 5, visando um *average dwell-time* que torne o NCS estável. Os parâmetros a e b, além de serem ajustados visando a estabilidade, também devem ser ajustados para dificultar o ataque de identificação de sistemas. Sob este aspecto, a literatura [Baştuğ 2012] indica que pequenos *dwell times* dificultam o processo de identificação do sistema.

Assim, no que diz respeito ao Objetivo I, especificamente por questão de estabilidade, a e b são aumentados, tanto quanto possível, para garantir o *average dwell-time* necessário para a estabilidade. Por outro lado, no que diz respeito ao Objetivo II – *i.e.* dificultar o ataque de identificação de sistemas –, a e b devem ser diminuídos, tanto quanto possível, para reduzir os *dwell times* do SLS. Considerando estes aspectos, no presente trabalho, a e b são empiricamente ajustados para satisfazer os dois objetivos conflitantes.

É digno de nota que, além de dificultar o Ataque Ativo de Identificação de Sistemas proposto em [de Sá et al. 2017a], a regra de chaveamento aleatório adotada também dificulta o lançamento de ataques furtivos/dependentes de modelo – mesmo que o atacante conheça as funções de controle $C_i(z)$. Isto decorre do simples fato de que é mais difícil sincronizar o ataque furtivo/dependente de modelo com as funções de controle $C_i(z)$ que alternam em intervalos aleatórios.

5. Resultados

Nesta seção, a performance da contramedida proposta é analisada em face do Ataque Ativo de Identificação de Sistemas de [de Sá et al. 2017a]. Nas simulações são feitas comparações entre dois NCSs: um com a contramedida proposta – *i.e.* utilizando um controlador chaveado; e outro sem a contramedida proposta – *i.e.* utilizando um controlador não chaveado. A performance do ataque a cada um dos NCSs é avaliada por meio de um conjunto de simulações realizadas no MATLAB. Os modelos dos NCSs, bem como os parâmetros do Ataque Ativo de Identificação de Sistemas, são descritos na Seção 5.1.

Cabe lembrar que, conforme definido na Seção 4, o projeto do controlador chaveado deve atender simultaneamente dois objetivos: cumprir os requisitos de controle da planta; e dificultar o processo de identificação. Neste sentido, a Seção 5.2 analisa os resultados do controlador chaveado sob o ponto de vista de uma contramedida para o Ataque Ativo de Identificação de Sistemas. A Seção 5.3, por sua vez, analisa o desempenho da contramedida proposta sob o ponto de vista do controle, a fim de identificar possíveis relações de compromisso que possam existir entre os dois objetivos.

5.1. NCSs Atacados e Parâmetros do Ataque

O NCS sem a contramedida proposta – também referido neste artigo como *sistema com modelo vulnerável* – é o mesmo NCS atacado em [de Sá et al. 2017a]. Consiste em

um motor DC cuja velocidade de rotação é controlada por um controlador Proporcional-Integral (PI), não chaveado. A função de transferência do motor DC $P(z)$ e a função de controle PI $C_1(z)$ são representadas por (5) e (6), respectivamente:

$$P(z) = \frac{0.3379z + 0.2793}{z^2 - 1.5462z + 0.5646}, \quad (5) \qquad C_1(z) = \frac{0.1701z - 0.1673}{z - 1}. \quad (6)$$

Assim, a função de transferência em malha fechada do *sistema com modelo vulnerável* $G_1(z)$, a ser identificada pelo atacante, é definida por (7):

$$G_1(z) = C_1(z)P(z) = \frac{g_{1,1}z^2 + g_{2,1}z + g_{3,1}}{z^3 + g_{4,1}z^2 + g_{5,1}z + g_{6,1}}, \quad (7)$$

onde $g_{1,1} = 0.0575$, $g_{2,1} = -0.0090$, $g_{3,1} = -0.0467$, $g_{4,1} = -2.5462$, $g_{5,1} = 2.1108$ e $g_{6,1} = -0.5646$.

O NCS dotado da contramedida proposta – *i.e.* com controlador chaveado – também controla um motor DC definido pela função de transferência (5). O controlador chaveado alterna entre duas funções de controle: $C_1(z)$, que é a mesma função de controle (6) do *sistema com modelo vulnerável*; e $C_2(z)$ que é definida por (8):

$$C_2(z) = \frac{0.1208z - 0,1167}{z - 1}. \quad (8)$$

Portanto, o NCS com o controlador chaveado é um SLS composto por dois subsistemas, cada um tendo uma função de transferência em malha aberta. As duas funções de transferência em malha aberta são, respectivamente: $G_1(z)$, que é a mesma função (7) do *sistema com modelo vulnerável*; e $G_2(z)$, definida por (9):

$$G_2(z) = C_2(z)P(z) = \frac{g_{1,2}z^2 + g_{2,2}z + g_{3,2}}{z^3 + g_{4,2}z^2 + g_{5,2}z + g_{6,2}}, \quad (9)$$

onde $g_{1,2} = 0.0408$, $g_{2,2} = -0.0057$, $g_{3,2} = -0.0326$, $g_{4,2} = -2.5462$, $g_{5,2} = 2.1108$ e $g_{6,2} = -0.5646$. Note que os denominadores de $G_1(z)$ e $G_2(z)$ são iguais, visto que apenas os numeradores de $C_1(z)$ e $C_2(z)$ são diferentes. Portanto, $g_{4,1} = g_{4,2}$, $g_{5,1} = g_{5,2}$ e $g_{6,1} = g_{6,2}$.

As funções de controle $C_1(z)$ e $C_2(z)$ são projetadas de forma que os dois subsistemas deste SLS sejam estáveis. Conforme descrito na Seção 4, as funções de controle são chaveadas aleatoriamente com base na cadeia de Markov apresentada na Figura 4, sob uma regra de comutação restrita, cujas restrições são estabelecidas pela FDP apresentada na Figura 5. Os parâmetros a e b da FDP foram empiricamente ajustados para $a = 20$ e $b = 40$, a fim de atender aos Objetivos I e II, discutidos na Seção 5. Cabe ressaltar que, em relação ao Objetivo I, os parâmetros a e b foram empiricamente ajustados visando, em primeiro lugar, a estabilidade global do sistema. No entanto, o tempo de acomodação e o *overshoot* também são avaliados nestas simulações.

O ataque é implementado utilizando o BSA, tendo em vista que esta metaheurística obteve o melhor desempenho nas simulações de ataque de [de Sá et al. 2017a]. Os parâmetros do BSA são os mesmos utilizados em [de Sá et al. 2017a]: a população possui 100 indivíduos; os limites de cada dimensão do espaço de busca são $[-10, 10]$; e η – que estabelece a amplitude do deslocamento dos indivíduos do BSA – é ajustado para 1. Em cada simulação, o BSA é executado por 4500 iterações.

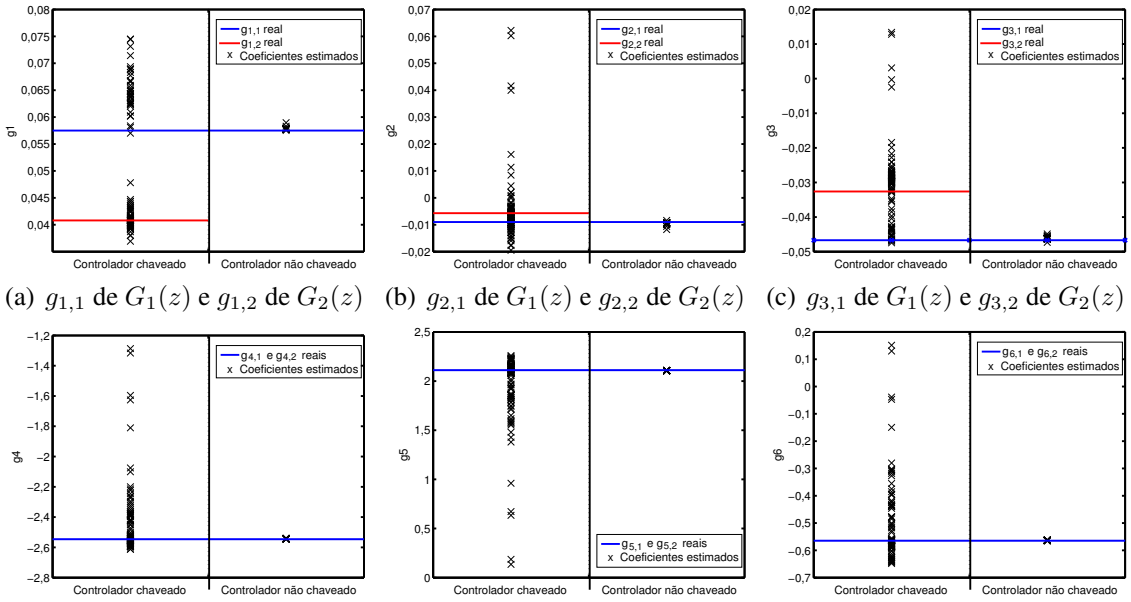
Assim como em [de Sá et al. 2017a], o sinal de ataque $a(k)$ mostrado na Figura 2 é um impulso unitário (10):

$$a(k) = \begin{cases} 1 & \text{se } k = k_a; \\ 0 & \text{se } k \neq k_a, \end{cases} \tag{10}$$

onde k_a é a única amostra em que o o atacante interfere no sistema, adicionando 1 ao sinal de realimentação. Em cada simulação, o sinal de realimentação é capturado pelo atacante durante um período $T = 2s$ (100 amostras), iniciando na amostra $k_a + 1$. Nos dois NCSs, a taxa de amostragem é de 50 amostras/s e a entrada $r(k)$ é uma degrau unitário. As simulações deste artigo não consideram perda de pacotes nem atrasos na rede.

5.2. Desempenho como Contramedida

Nesta seção, são apresentados os resultados obtidos pelo ataque Ativo de Identificação de Sistemas, quando lançado sobre os NCSs descritos na Seção 5.1 – um NCS utilizando o controlador chaveado e o outro utilizando o controlador não chaveado. Em cada NCS foram executadas 100 simulações de ataque.



(a) $g_{1,1}$ de $G_1(z)$ e $g_{1,2}$ de $G_2(z)$ (b) $g_{2,1}$ de $G_1(z)$ e $g_{2,2}$ de $G_2(z)$ (c) $g_{3,1}$ de $G_1(z)$ e $g_{3,2}$ de $G_2(z)$

(d) $g_{4,1}$ de $G_1(z)$ e $g_{4,2}$ de $G_2(z)$ (e) $g_{5,1}$ de $G_1(z)$ e $g_{5,2}$ de $G_2(z)$ (f) $g_{6,1}$ de $G_1(z)$ e $g_{6,2}$ de $G_2(z)$

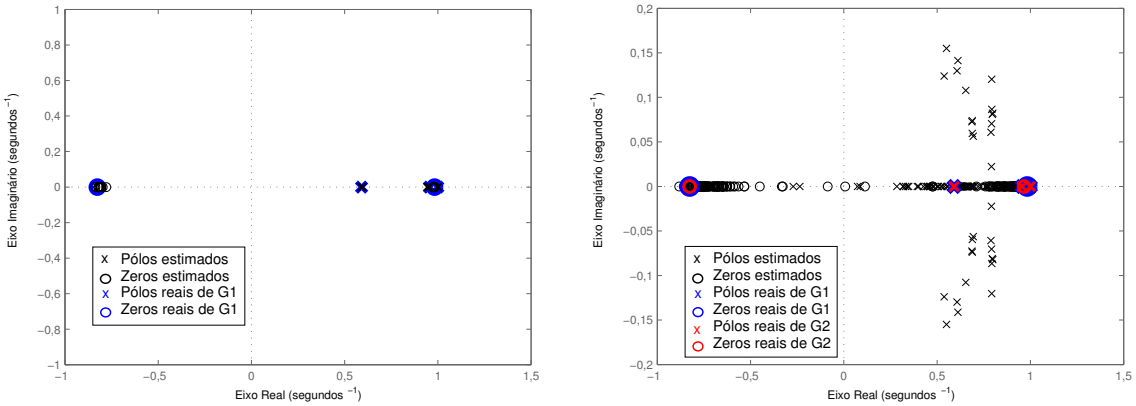
Figura 6. Coeficientes estimados pelo ataque no NCS com a contramedida proposta (*i.e.* com controlador chaveado) e no NCS sem a contramedida proposta (*i.e.* com controlador não chaveado).

Todos os coeficientes estimados nas 100 simulações de ataque em cada NCS são apresentados na Figura 6. Cabe lembrar que o NCS que utiliza o controlador não chaveado só possui uma função de transferência em malha aberta $G_1(z)$, enquanto que o NCS com o controlador chaveado possui duas funções $G_1(z)$ and $G_2(z)$. Note que os valores reais dos coeficientes $[g_{1,1}, g_{2,1}, g_{3,1}, g_{4,1}, g_{5,1}, g_{6,1}]$ e $[g_{1,2}, g_{2,2}, g_{3,2}, g_{4,2}, g_{5,2}, g_{6,2}]$ das funções $G_1(z)$ e $G_2(z)$, respectivamente, também são representados na Figura 6. Observando as Figuras 6(a) a 6(f), é possível verificar que os coeficientes estimados no NCS com o controlador não chaveado são precisos e exatos. Neste NCS, o ataque Ativo de Identificação de Sistemas fornece a informação e a confiança que o atacante necessita para projetar ataques furtivos/dependentes de modelo. Por outro lado, no NCS dotado da contramedida proposta, o uso do controlador chaveado causa a dispersão dos valores

estimados, reduzindo a precisão e a exatidão dos coeficientes obtidos pelo atacante. Conforme apresentado na Figura 6, os valores estimados neste SLS são difusos e não indicam, de forma acurada, nenhum dos coeficientes de $G_1(z)$ e $G_2(z)$.

O impacto do controlador chaveado na performance do ataque também pode ser verificado ao comparar os valores mínimos globais encontrados, pelo BSA, para a função de aptidão (4). No NCS dotado com o controlador chaveado, os valores mínimos globais de todas as simulações de ataque estão entre $1,81 \times 10^{-06}$ e $1,96 \times 10^{-04}$ (a média é $2,50 \times 10^{-05}$ e o desvio padrão é $3,97 \times 10^{-05}$). Por outro lado, no NCS com o controlador não chaveado, todos os valores mínimos globais estão entre $7,82 \times 10^{-09}$ e $4,46 \times 10^{-08}$ (a média é $8,75 \times 10^{-09}$ e o desvio padrão é $4,80 \times 10^{-09}$). Lembre-se que, conforme discutido na Seção 3, sem perturbações ou ruído, o valor mínimo de (4) é $\min f_j = 0$ quando o sistema atacado é identificado de forma perfeita. Dessa forma, a maior ordem dos valores mínimo globais, causada pelo uso do controlador chaveado, também demonstra a efetividade da contramedida proposta. Do ponto de vista do atacante, estes valores mínimos globais majorados podem ser um indicativo de que o Ataque Ativo de Identificação de Sistemas não foi efetivo em obter o modelo do sistema atacado. Neste caso, o atacante deve hesitar em lançar um ataque furtivo/dependente de modelo com base nas informações obtidas pelo Ataque Ativo de Identificação de Sistemas.

O impacto da contramedida no referido ataque também pode ser verificado nos diagramas de pólos e zeros da Figura 7. Na Figura 7(a), são apresentados os pólos e zeros das funções de transferência em malha aberta estimadas nas 100 simulações com o controlador não chaveado. A Figura 7(b), por sua vez, apresenta os pólos e os zeros das funções de transferência em malha aberta estimadas nas simulações utilizando o controlador chaveado. Note que, nas simulações com o controlador não chaveado, os pólos e zeros estimados convergem, de forma acurada, para os pólos e zeros da função de transferência em malha aberta $G_1(z)$. Por outro lado, a Figura 7(b) mostra que, quando a contramedida proposta é utilizada, os pólos e zeros estimados se espalham e não concorrem para os pólos e zeros reais de $G_1(z)$ e $G_2(z)$ – as funções de transferência em malha aberta dos dois subsistemas do SLS.



(a) NCS sem a contramedida proposta (b) NCS com a contramedida proposta

Figura 7. Pólos e zeros obtidos com o Ataque Ativo de Identificação de Sistemas.

O espalhamento dos pólos e zeros da Figura 7(b), a imprecisão dos coeficientes apresentados na Figura 6, e os elevados mínimos globais encontrados pelo BSA demonstram a efetividade do uso de controladores chaveados como uma contramedida para o Ataque Ativo de Identificação de Sistemas de [de Sá et al. 2017a]. Com a contramedida

proposta, é possível afirmar que o modelo obtido é impreciso/ambíguo, de tal forma que o atacante hesite em lançar contra o sistema ataques que sejam furtivos/dependentes de modelo. Assim, o Objetivo II definido na Seção 4 é atendido.

5.3. Desempenho como Controlador

Nesta seção, o desempenho da contramedida proposta é analisado do ponto de vista do controle, a fim de identificar possíveis impactos da mesma no controle da planta. Para isso, os seguintes aspectos são avaliados: estabilidade; *overshoot*; e tempo de acomodação. Com base nestes aspectos, o desempenho do controlador chaveado é comparado com o desempenho do controlador não chaveado. Tendo em vista a natureza estocástica do controlador chaveado descrito na Seção 5.1, o qual chaveia aleatoriamente entre duas funções de controle, os aspectos mencionados são avaliados por meio de um conjunto de 100.000 simulações.

Na Figura 8, são apresentadas as respostas de ambos os NCSs no domínio do tempo. As respostas do NCS dotado da contramedida proposta são representadas pela área destacada. Os limites desta área são desenhados com base nos valores máximos e mínimos da saída da planta, considerando todas as 100.000 simulações. Em outras palavras, com o controlador chaveado, todos os sinais de saída $y(k)$ gerados nas simulações estão contidos na área destacada. A resposta, não estocástica, do NCS com o controlador não chaveado é representada pela linha vermelha da Figura 8. Note que, até $t = 0.4s$ as respostas obtidas com o controlador chaveado são iguais à resposta obtida com o controlador não chaveado. Isto é causado pelo mínimo *dwell time* de $0.4s$, que corresponde à quantidade mínima de intervalos de amostragem que o sistema deve permanecer no mesmo estado (estabelecida na Seção 5.1 como $a = 20$ amostras). Com base na Figura 8, considerando todas as 100.000 simulações, é possível verificar que o NCS com a contramedida proposta é estável, a saída da planta converge para o *set point* ($r(k) = 1rad/s$) sem erro estacionário, e a planta não apresenta *overshoots*. Considerando estes aspectos, do ponto de vista do controle, a contramedida proposta apresenta o mesmo desempenho que o controlador não chaveado.

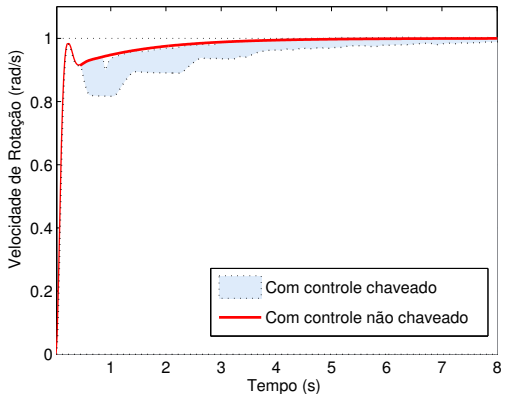


Figura 8. Resposta do motor no domínio do tempo.

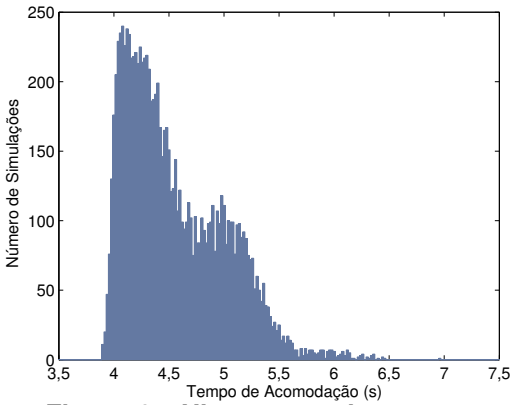


Figura 9. Histograma dos tempos de acomodação quando a contramedida proposta é utilizada.

Por outro lado, devido aos chaveamentos sucessivos, é possível verificar na Figura 8 que os tempos de acomodação obtidos com a contramedida proposta são maiores do que o tempo de acomodação obtido com o controlador não chaveado. O tempo de acomodação, determinístico, do NCS com o controlador não chaveado é $2,4s$. Já o tempo de acomodação t_s obtido com o controlador chaveado é estocástico e depende da

sequência de *dwell times* ocorridos antes de alcançar t_s, o que é aleatório. Os tempos de acomodação de todas as 100.000 simulações com o controlador chaveado são apresentados no histograma da Figura 9. Os tempos de acomodação máximo e mínimo são $3,90s$ e $6,96s$, respectivamente. A média é $4,555 \pm 0,0088s$, com intervalo de confiança de 95%.

O desempenho da contramedida proposta, na perspectiva do controle, é satisfatório e indica a viabilidade de atender aos Objetivos I e II, simultaneamente. Nestas simulações, o controle realizado pelo controlador chaveado apresenta uma performance similar à do controlador não chaveado. O requisito primário do Objetivo I – *i.e.* a estabilidade – é satisfeito, bem como o requisito de não causar *overshoot* na planta. No entanto, as simulações indicam um aumento no tempo de acomodação do sistema, o que pode não ser crítico, mas deve ser analisado dependendo do processo específico que está sendo controlado. Sendo assim, ao decidir pelo uso desta contramedida, deve-se levar em conta a relação de compromisso que existe entre mitigar o ataque de identificação e aumentar o tempo de acomodação do sistema.

6. Conclusão

Neste artigo, é proposto o uso de um controlador aleatoriamente chaveado como contramedida para o Ataque Ativo de Identificação de Sistemas, em caso de falha das contramedidas convencionais – tais como o uso de criptografia e arquiteturas de segurança de rede. É demonstrado, por meio de simulações, que esta contramedida é capaz de mitigar o ataque mencionado, tornando o modelo obtido pelo atacante impreciso/ambíguo. Ao mesmo tempo, as simulações demonstram que o desempenho da contramedida proposta é satisfatório do ponto de vista do controle. Considerando os aspectos de controle, em geral, a contramedida proposta apresenta uma performance similar à performance do controlador não chaveado, com um aumento do tempo de acomodação – o que, em muitas aplicações não representa um óbice.

Como trabalhos futuros, planejamos avaliar o desempenho desta contramedida em face de outros ataques/algoritmos de identificação de sistemas. Encorajamos também o desenvolvimento de uma heurística ou um método analítico capaz de prover funções de controle e regras de chaveamento que maximizem o desempenho da contramedida, no que concerne aos dois objetivos apresentados – *i.e.* cumprir os requisitos de controle da planta; e dificultar o processo de identificação.

Agradecimentos

Esta pesquisa foi parcialmente apoiada pelos órgãos de fomento CNPq e FAPERJ.

Referências

Amin, S., Litrico, X., Sastry, S., and Bayen, A. M. (2013). Cyber security of water scada systems part i: analysis and experimentation of stealthy deception attacks. *IEEE Transactions on Control Systems Technology*, 21(5):1963–1970.

Başţuğ, M. (2012). Recursive modeling of switched linear systems: a behavioral approach. Master's thesis, Istambul Technical University.

Civicioglu, P. (2013). Backtracking search optimization algorithm for numerical optimization problems. *Applied Mathematics and Computation*, 219(15):8121–8144.

Dasgupta, S., Routh, A., Banerjee, S., Agilageswari, K., Balasubramanian, R., Bhandarkar, S., Chattopadhyay, S., Kumar, M., and Gupta, A. (2013). Networked control

of a large pressurized heavy water reactor (phwr) with discrete proportional-integral-derivative (pid) controllers. *IEEE Transactions on Nuclear Science*, 60(5):3879–3888.

de Sá, A. O., da Costa Carmo, L. F. R., and Machado, R. C. S. (2017a). Bio-inspired active attack for identification of networked control systems. In *10th EAI International Conference on Bio-inspired Information and Communications Technologies (BICT)*, pages 1–8. ACM.

de Sá, A. O., da Costa Carmo, L. F. R., and Machado, R. C. S. (2017b). Covert attacks in cyber-physical control systems. *IEEE Transactions on Industrial Informatics*, 13(4):1641–1651.

Hespanha, J. P. and Morse, A. S. (1999). Stability of switched systems with average dwell-time. In *Decision and Control, 1999. Proceedings of the 38th IEEE Conference on*, volume 3, pages 2655–2660. IEEE.

Kennedy, J. e Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of 1995 IEEE International Conference on Neural Networks*, pages 1942–1948.

Lin, H. and Antsaklis, P. J. (2009). Stability and stabilizability of switched linear systems: a survey of recent results. *IEEE Transactions on Automatic control*, 54(2):308–322.

Morse, A. S. (1996). Supervisory control of families of linear set-point controllers-part i. exact matching. *IEEE Transactions on Automatic Control*, 41(10):1413–1431.

Pang, Z.-H. and Liu, G.-P. (2012). Design and implementation of secure networked predictive control systems under deception attacks. *IEEE Transactions on Control Systems Technology*, 20(5):1334–1342.

Skafidas, E., Evans, R. J., Savkin, A. V., and Petersen, I. R. (1999). Stability results for switched controller systems. *Automatica*, 35(4):553–564.

Smith, R. (2011). A decoupled feedback structure for covertly appropriating networked control systems. In *Proceedings of the 18th IFAC World Congress 2011*, volume 18. IFAC-PapersOnLine.

Smith, R. S. (2015). Covert misappropriation of networked control systems: Presenting a feedback structure. *Control Systems, IEEE*, 35(1):82–92.

Stouffer, K., Pillitteri, V., Lightman, S., Abrams, M., and Hahn, A. (2015). Nist special publication 800-82, revision 2: Guide to industrial control systems (ics) security. *Gaithersburg, MD, USA: National Institute of Standards and Technology*.

Teixeira, A., Shames, I., Sandberg, H., and Johansson, K. H. (2015). A secure control framework for resource-limited adversaries. *Automatica*, 51:135–148.

Wang, J. (2013). *Identification of Switched Linear Systems*. PhD thesis, University of Alberta.

Yang, C., Guan, Z.-H., and Huang, J. (2011). Stochastic switched controller design of networked control systems with a random long delay. *Asian Journal of Control*, 13(2):255–264.

Zhai, G., Hu, B., Yasuda, K., and Michel, A. N. (2002). Qualitative analysis of discrete-time switched systems. In *American Control Conference, 2002. Proceedings of the 2002*, volume 3, pages 1880–1885. IEEE.

