SCIENCE CHINA
Information Sciences

• LETTER •

September 2023, Vol. 66 199204:1–199204:2
https://doi.org/10.1007/s11432-021-3644-3

Attack detectability and stealthiness in distributed optimal coordination of cyber-physical systems

Liwei AN[1] & Guang-Hong YANG[1,2*]

[1]*College of Information Science and Engineering, Northeastern University, Shenyang* 110819*, China;*
[2]*State Key Laboratory of Synthetical Automation for Process Industries (Northeastern University), Shenyang* 110819*, China*

Received 12 December 2021/Revised 9 July 2022/Accepted 22 November 2022/Published online 21 April 2023

Citation An L W, Yang G-H. Attack detectability and stealthiness in distributed optimal coordination of cyber-physical systems. Sci China Inf Sci, 2023, 66(9): 199204, https://doi.org/10.1007/s11432-021-3644-3

With potential applications of distributed optimization (DO) in large-scale cyber-physical systems (CPSs), many important results on distributed optimal coordination (DOC) algorithms have been reported for multi-agent systems with various physical dynamics, which have wide applications such as the cooperative search of radio sources, the motion coordination, and the distributed optimal power flow (see [1] and references therein). Given the growing threat of malicious attacks in large-scale (and security-critical) CPSs, the vulnerability study of consensus-based DOC algorithms becomes an important issue. Some recent studies (e.g., [2] and references therein) considered the problem of resilient DO under adversarial models, however, without considering the agent's own physical dynamics.

In this study, we investigate the attack detection and identification (ADI) in the DOC of a group of nonlinear CPSs in the presence of cyber attacks by extending the existing residual-based fault detection and isolation (FDI) methods [3–5]. However, the existing FDI results mainly focus on the design of detection and identification mechanism but do not consider the detectability and vulnerability for malicious attacks that aim to destroy the system function without being detected. This is the focus problem to be concerned with in this study. We make two modifications for the existing residual-based FDI methods: (1) double coupling residuals are generated by a distributed filter; (2) "strongly-robust" thresholds with prescribed performance are designed to enhance the identifiability. The main contribution of this study is to provide an extensive analysis for attack detectability and stealthiness and prove that (1) local detectable attacks on a subsystem can be exactly identified from multiple attacks propagated from its neighbors (Theorem 1) and (2) undetectable attacks cannot destroy the system convergence (Theorem 2). Note that some effective distributed FDI and ADI methods for interconnected systems have also been proposed in the existing literature. Detailed related studies and comparisons can be seen in Appendix A.

DOC architecture. Consider a CPS consisting of N subsystems, which aims at achieving the DOC task. The jth subsystem, $j = 1,\dots,N$, is described by the pair $(\mathcal{P}^{(j)},\mathcal{C}^{(j)})$, where $\mathcal{C}^{(j)}$ denotes the cyber part which is responsible for task decision-making, while $\mathcal{P}^{(j)}$ denotes the physical part which is responsible for task execution. The physical part $\mathcal{P}^{(j)}$ is modeled as a nonlinear system:

$$\Sigma^{(j)}:\begin{cases}\dot{x}_i^{(j)}(t)=x_{i+1}^{(j)}(t)+\varphi_i^{(j)}(\bar{x}_i^{(j)}(t))\theta_j,\\ \dot{x}_n^{(j)}(t)=\beta_j u^{(j)}(t)+\varphi_n^{(j)}(x^{(j)}(t))\theta_j,\\ y^{(j)}(t)=x_1^{(j)}(t)+a^{(j)}(t),\end{cases}\tag{1}$$

where $i=1,\dots,n-1$ $(n>1)$, $x_i^{(j)}(t)\in\mathbb{R}^m$, $\bar{x}_i^{(j)}(t)=\mathrm{vec}(x_1^{(j)}(t),\dots,x_i^{(j)}(t))\in\mathbb{R}^{im}$, $x^{(j)}(t)=\bar{x}_n^{(j)}(t)\in\mathbb{R}^{nm}$ is the state variable; $u^{(j)}(t)\in\mathbb{R}^m$ is the control input; $\varphi_i^{(j)}(\bar{x}^{(j)}(t))\in\mathbb{R}^{m\times p}$ and $\beta_j\in\mathbb{R}^{m\times m}$ are known nonlinear and smooth function matrices and β_j is nonsingular; $\theta_j\in\mathbb{R}^p$ is an unknown constant vector; $y^{(j)}(t)\in\mathbb{R}^m$ is the output measurement transmitted to the cyber superstratum through a wireless network channel which may be corrupted by the attack signal $a^{(j)}(t)\in\mathbb{R}^m$. In particular, to provide security guarantees against worst case adversarial behavior in theory, we allow the adversarial attacker to know the overall system model, system state, control input, and the possible fault detector $\mathfrak{D}$ (e.g., distributed adaptive observers [4,5]) equipped on the CPS. Thus, the attack signal can be modeled as $a^{(j)}(t)=\phi^{(j)}(x(t),u(t),\mathfrak{D},t-T_a^{(j)})$ where $\phi^{(j)}(\cdot,\cdot,\cdot,\cdot)\in\mathbb{R}^m$ is an unknown function. Due to the adversary's strategic design, here we assume that $a^{(j)}(t)$ in system (1) denotes a strategic attack model which can find and exploit the vulnerability of $\mathfrak{D}$ to destroy the function of the system without being detected (see related design methods of stealthy attacks against various fault detectors, e.g., [6] and references therein). The overall DOC architecture can be found in Appendix B.

The objective of the CPS architecture is to steer all the physical subsystems to cooperatively reach the optimal output that minimizes the team performance function:

$$\min\sum_{j=1}^{N}g^{(j)}(\hbar),\ \hbar\in\mathbb{R}^m,\tag{2}$$

* Corresponding author (email: yangguanghong@ise.neu.edu.cn)

where $g^{(j)}:\mathbb{R}^m\to\mathbb{R}$ is a convex and differentiable performance function privately known to the agent $\mathcal{D}^{(j)}$. The optimization module $\mathcal{O}^{(j)}$ and control agent $\mathcal{K}^{(j)}$ form a basic DOC scheme under healthy conditions, i.e., $a^{(j)}(t)=0$ for any $t\geqslant 0$ and $j\in\{1,\ldots,N\}$. Based on the existing DO algorithm [7], the model of the optimization module $\mathcal{O}^{(j)}$ is designed as the following algorithm:

$$\mathcal{O}^{(j)}:\begin{cases}\dot{y}_r^{(j)}=-\nabla g^{(j)}(y_r^{(j)})-\tilde{v}^{N_j}\\ \qquad\quad -(1+\eta)\sum\limits_{i\in N_j}w_{ji}(y^{(j)}-y^{(i)}),\\ \dot{v}^{(j)}=\sum\limits_{i\in N_j}w_{ji}(y^{(j)}-y^{(i)}),\end{cases}\tag{3}$$

where $\tilde{v}^{N_j}=\sum_{i\in N_j}w_{ji}(v^{(j)}-v^{(i)})\in\mathbb{R}^m$; $y_r^{(j)}\in\mathbb{R}^m$ and $v^{(j)}\in\mathbb{R}^m$ are state vectors; $\nabla g^{(j)}$ is the gradient of $g^{(j)}$; $\eta>0$ is a parameter; N_j is the set of neighbors of node j; $w_{ji}>0$ is the weight. Applying the standard adaptive backstepping control with prescribed performance [8] to system (1), we derive the following controller:

$$u^{(j)}=\mathcal{K}^{(j)}\left(y_r^{(j)},x^{(j)},S\left(\frac{z_1^{(j)}}{\delta^{(j)}}\right)\right),\tag{4}$$

$$S\left(\frac{z_1^{(j)}}{\delta^{(j)}}\right)=\frac{1}{2}\ln\left(1+\frac{z_1^{(j)}}{\delta^{(j)}}\right)-\frac{1}{2}\ln\left(1-\frac{z_1^{(j)}}{\delta^{(j)}}\right),$$

where $z_1^{(j)}=x_1^{(j)}-y_r^{(j)}$, $\delta^{(j)}(t)$ is a prescribed performance bound function such that $|z_{1,s}^{(j)}(0)|<\delta^{(j)}(0)$, and $z_{1,s}^{(j)}$ denotes the sth ($s=1,\ldots,m$) element of $z_1^{(j)}$. Under the assumption that the network topology $\mathcal{G}$ is connected, the closed-loop CPS $(\mathcal{P}^{(j)},\mathcal{C}^{(j)})$ with $(\mathcal{K}^{(j)},\mathcal{O}^{(j)})$, $j=1,\ldots,N$ achieves output consensus at an optimal solution of problem (2). Also, from [8], one has $z_1^{(j)}(t)\in\Delta_z^{\delta^{(j)}}$ where $\Delta_z^{\delta^{(j)}}:=\{z(t)\in\mathbb{C}_m^n:\int_{\tau=0}^t\|z(\tau)\|^2\mathrm{d}\tau\leqslant\Omega,\|z(t)\|\leqslant\sqrt{m}\delta^{(j)}(t)\}$ in the absence of cyber attacks, Ω is a known constant determined by the initial state, the upper bound of $\theta_i^{(j)}$ (determined by physical laws), and controller parameters [8]. The detailed controller design and stability analysis can be found in Appendix C.

ADI design and analysis. The objective of the ADI is to design the monitoring module $\mathcal{M}^{(j)}$ which can detect and identify the local attack $a^{(j)}$, $j=1,\ldots,N$. The ADI structure follows the standard framework of residual-based FDI [3–5], consisting of detection filters, residuals, thresholds, and decision logic. The detailed design procedure of the ADI mechanism can be found in Appendix D.

• Detection filter

$$\mathcal{M}^{(j)}:\begin{cases}\dot{\hat{y}}_r^{(j)}=-\nabla g^{(j)}(y_r^{(j)})-\tilde{v}^{N_j}\\ \qquad\quad -(1+\eta)\sum\limits_{i\in N_j}w_{ji}(\hat{y}_r^{(j)}-y^{(i)}),\\ \dot{\hat{v}}^{(j)}=\sum\limits_{i\in N_j}w_{ji}\left[(\hat{y}_r^{(j)}-y^{(i)})-(v^{(j)}-\hat{v}^{(j)})\right],\end{cases}\tag{5}$$

where $\hat{y}_r^{(j)}\in\mathbb{R}^m$ and $\hat{v}^{(j)}\in\mathbb{R}^m$ are the estimates of $y_r^{(j)}$ and $v^{(j)}$ (even $y_r^{(j)}$ and $v^{(j)}$ are available for $\mathcal{M}^{(j)}$), respectively, based on the local communication signals $y^{(i)}$ and $v^{(i)}$, $i\in\{j\}\cup N_j$.

• Double residuals

$$e_r^{(j)}=y_r^{(j)}-\hat{y}_r^{(j)},\quad e_v^{(j)}=v^{(j)}-\hat{v}^{(j)}.\tag{6}$$

• "Strongly-robust" thresholds

$$\bar{e}_{r,H}^{(j)}(t)=e^{-\eta^{(j)}t}e_{r,H}^{(j)}(0)+\bar{\Psi}^{(j)}_{\Delta_z^{\delta^{(j)}}}(\eta^{(j)},0,t),$$

$$\bar{e}_{v,H}^{(j)}(t)=e^{-w_{N_j}t}e_{v,H}^{(j)}(0)+\bar{\Psi}^{(j)}_{\Delta_{ez}^{\delta^{(j)}}}(w_{N_j},0,t),\tag{7}$$

where $\Delta_{ez}^{\delta^{(j)}}:=\{e+z:\|e\|\leqslant\bar{e}_{r,H}^{(j)},\ z\in\Delta_z^{\delta^{(j)}}\}$ and $\bar{\Psi}^{(j)}_{\Delta^{\delta^{(j)}}}(\alpha,t_0,t):=\sup_{h(t)\in\Delta^{\delta^{(j)}}}\alpha\int_{\tau=t_0}^t e^{\alpha(\tau-t)}\|h(\tau)\|\mathrm{d}\tau$.

• Decision logic

$$\mho^{(j)}(t)=\mho^{(j,r)}(t)\cup\mho^{(j,v)}(t),\tag{8}$$

where $\mho^{(j,r)}(t):\ \|e_r^{(j)}(t)\|\leqslant\bar{e}_{r,H}^{(j)}(t)$ and $\mho^{(j,v)}(t):\|e_v^{(j)}(t)\|\leqslant\bar{e}_{v,H}^{(j)}(t)$. If $\mho^{(j)}(t)$ is violated, then $\mathcal{M}^{(j)}$ will generate an alarm.

The following two theorems give the analysis of the attack detection and stealthiness (i.e., vulnerability analysis).

Theorem 1. Consider the ADI mechanism defined in (5)–(8). If there is a time instant $T_d^{(j)}$ when $\mho^{(j)}(T_d^{(j)})$ is violated and $\int_{t=0}^{T_d^{(j)}}\delta^{(j)2}(t)\mathrm{d}t\leqslant\Omega/m$, then the occurrence of local attack $a^{(j)}$ is guaranteed.

The proof and remark can be found in Appendix E.

Theorem 2. The closed-loop CPS $(\mathcal{P}^{(j)},\mathcal{C}^{(j)})$ with $(\mathcal{K}^{(j)},\mathcal{D}^{(j)}(\mathcal{O}^{(j)},\ \mathcal{M}^{(j)}))$ achieves output consensus at an optimal solution of problem (2) even in the presence of the undetectable attacks.

The proof can be found in Appendix F.

From Theorems 1 and 2, we can draw a conclusion: undetectable attacks cannot destroy the system convergence, while locally detectable attacks can be exactly identified even in the presence of attack's coupling impacts caused by the communications among subsystems.

The simulation illustration can be found in Appendix G.

Acknowledgements This work was supported by National Key Research and Development Program of China (Grant No. 2020YFE0201100), National Natural Science Foundation of China (Grant Nos. 61621004, 62103089, U1908213), and Research Fund of State Key Laboratory of Synthetical Automation for Process Industries, China (Grant No. 2018ZCX03).

Supporting information Appendixes A–G. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 An L, Yang G-H. Distributed optimal coordination for heterogeneous linear multiagent systems. IEEE Trans Automat Contr, 2022, 67: 6850–6857
2 Zhao C, He J, Wang Q-G. Resilient distributed optimization algorithm against adversarial attacks. IEEE Trans Automat Contr, 2020, 65: 4308–4315
3 Shames I, Teixeira A M H, Sandberg H, et al. Distributed fault detection for interconnected second-order systems. Automatica, 2021, 47: 2757–2764
4 Zhang Q, Zhang X. Distributed sensor fault diagnosis in a class of interconnected nonlinear uncertain systems. Annu Rev Control, 2013, 37: 170–179
5 Reppa V, Polycarpou M M, Panayiotou C G. Distributed sensor fault diagnosis for a network of interconnected cyberphysical systems. IEEE Trans Control Netw Syst, 2015, 2: 11–23
6 Teixeira A, Shames I, Sandberg H, et al. A secure control framework for resource-limited adversaries. Automatica, 2015, 51: 135–148
7 Gharesifard B, Cortes J. Distributed continuous-time convex optimization on weight-balanced digraphs. IEEE Trans Automat Contr, 2014, 59: 781–786
8 Wang W, Wen C. Adaptive actuator failure compensation control of uncertain nonlinear systems with guaranteed transient performance. Automatica, 2010, 46: 2082–2091