Automatica 173 (2025) 112071

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Privacy and security trade-off in interconnected systems with known or unknown privacy noise covariance[☆]

Haojun Wang[a], Kun Liu[a,*], Baojia Li[a], Emilia Fridman[b], Yuanqing Xia[a]

[a] School of Automation, Beijing Institute of Technology, Beijing, China
[b] School of Electrical Engineering, Tel Aviv University, Tel Aviv, Israel

ARTICLE INFO

Article history:
Received 1 January 2024
Received in revised form 26 September 2024
Accepted 10 October 2024
Available online xxxx

Keywords:
Privacy preservation
Attack detection
Trade-off
Interconnected systems

ABSTRACT

This paper is concerned with the security problem for interconnected systems, where each subsystem is required to detect local attacks. Moreover, we consider that there exists an additional eavesdropper being able to infer the private information. Then, a privacy-preserving method is employed by adding privacy noise to transmitted data, and the privacy level is measured by mutual information. Nevertheless, adding privacy noise to transmitted data may affect the detection performance metrics such as detection probability and false alarm probability. Thus, we theoretically analyze the trade-off between the privacy and the detection performance. An optimization problem with maximizing both the degree of privacy preservation and the detection probability is established to obtain the covariance of the privacy noise. In addition, the attack detector of each subsystem may not obtain all information about the privacy noise. We further theoretically analyze the trade-off between the privacy and the false alarm probability when the attack detector has no knowledge of the privacy noise covariance. An optimization problem with maximizing the degree of privacy preservation with guaranteeing a bound of false alarm distortion level is established to obtain the covariance of the privacy noise. Moreover, we consider that each subsystem can estimate the unknown privacy noise covariance by the secondary data. Based on the estimated covariance, we construct another attack detector and analyze how the privacy noise affects its detection performance. Finally, a numerical example is provided to verify the effectiveness of theoretical results.

1. Introduction

With the development of sensing, communication and control technology, large-scale systems have been applied in many fields, such as power systems (Liu, Chen, Zourntos, Kundur, & Butler-Purry, 2014), intelligent transportation (Dey, Mishra, & Chowdhury, 2014) and intelligent vehicles (Liu, Xu, & Ding, 2016). However, the communication networks in interconnected systems potentially suffer from the security and the privacy issues. Some malicious attackers attempt to compromise the integrity, availability, and confidentiality of data transmitted through communication networks, thereby deteriorating the systems performance and even leading disastrous consequences (Teixeira, Shames, Sandberg, & Johansson, 2015). Therefore, maintaining the security and the privacy becomes a key issue for interconnected systems.

For the security problems, to cope with the network attacks in interconnected systems, some distributed attack detectors have been developed, such as in Anguluri, Katewa, and Pasqualetti (2018), Boem, Gallo, Ferrari-Trecate, and Parisini (2017), Katewa, Anguluri, and Pasqualetti (2021), where each local detector only requires the locally available information and the knowledge of local model. Nevertheless, there may exist covert attacks to degrade the performance of each subsystem while keeping stealthy locally (Smith, 2015). Therefore in Barboni, Gallo, Boem, and Parisini (2019), Barboni, Rezaee, Boem, and Parisini (2020), to detect the covert attacks being stealthy, the distributed attack detectors are proposed based on the attack-sensitive-residuals by using the local information and the communicated estimates, where the detector of each subsystem can detect the covert attacks on its neighboring subsystems.

On the other hand, the information exchange between subsystems may leak private information. Therefore, it is necessary to consider the privacy-preserving methods. One mechanism is homomorphic cryptography (Lu & Zhu, 2018; Ruan, Gao, & Wang,

[☆] This work is supported by National Natural Science Foundation of China (no. U24A20264, 62273041), Open Projects of the Institute of Systems Science, Beijing Wuzi University (BWUISS11), and Israel Science Foundation (no. 446/24). The material in this paper was presented at None. This paper was recommended for publication in revised form by Associate Editor Tongwen Chen under the direction of Editor Florian Dorfler.

[*] Corresponding author.
E-mail addresses: haojunwang@bit.edu.cn (H. Wang), kunliubit@bit.edu.cn (K. Liu), jambooree@bit.edu.cn (B. Li), emilia@tauex.tau.ac.il (E. Fridman), xia_yuanqing@bit.edu.cn (Y. Xia).

https://doi.org/10.1016/j.automatica.2024.112071
0005-1098/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

2019), which can easily enable privacy preservation. However, it usually suffers from high communication burden and computation cost (He, Cai, Cheng, Pan, & Shi, 2018). Another mechanism is to add privacy noise to transmitted data. Differential privacy, which is realized by adding noises with Laplace or Gaussian distributions, has been applied in many fields, such as state estimation (Le Ny & Pappas, 2013), linear quadratic control (Yazdani, Jones, Leahy, & Hale, 2022) and distributed optimization (Han, Liu, Lin, & Xia, 2022). Moreover, there are methods to design privacy noise from an information-theoretic perspective, such as Fisher information (Farokhi & Sandberg, 2019), condition entropy (Nekouei, Skoglund, & Johansson, 2018), Kullback–Leibler divergence (Lin, Liu, Han, & Xia, 2024), and mutual information (Murguia, Shames, Farokhi, Nešić, & Poor, 2021).

Although adding privacy noise can improve the degree of privacy preservation, the attack detection performance such as false alarm probability and detection probability may be affected by the privacy noise. The trade-off between the privacy and the detection probability is analyzed for the single system in Farokhi and Esfahani (2018), where the privacy level is measured by Fisher information. Then the trade-off between the privacy and the detection probability is analyzed in Katewa et al. (2021) for interconnected systems, where the privacy level is measured by the estimation error covariance.

Moreover, it is noted that the above works are based on the condition that the attack detector has the knowledge of the privacy noise covariance. In order to further enhance privacy, the covariance of the privacy noise may be unknown to the attack detector, which means that the design of the attack detector cannot be based on the privacy noise. Then in Hayati, Murguia, and van de Wouw (2024), the trade-off between the privacy and the false alarm probability is analyzed for the signal system without knowing the privacy noise covariance, where the privacy level is measured by mutual information. Moreover, an optimization problem is established in Hayati et al. (2024) to obtain the optimal covariance of the privacy noise.

Inspired by the above discussion, in this paper we aim to analyze the trade-off between the privacy and the security for interconnected systems. The privacy-preserving method is to add privacy noise for keeping the state of each subsystem private, and the privacy is quantified by mutual information. The security is measured by the false alarm probability or the detection probability of the attack detectors. The main results are summarized as follows:

(1) We firstly analyze the trade-off between the privacy and the detection performance for the interconnected system. Moreover, an optimization problem based on mutual information is established for maximizing the degree of privacy preservation and the detection probability to obtain the covariance of the privacy noise.

(2) Then, when the privacy noise covariance is unknown to the attack detector, we not only theoretically analyze the trade-off between the privacy and the false alarm probability, but also establish an optimization problem to obtain the covariance of the privacy noise to maximize the privacy degree and guarantee a bound of false alarm distortion level.

(3) Furthermore, in order to analyze the effect of the privacy noise on the detection probability under the unknown privacy noise covariance, we consider that each subsystem can estimate the unknown privacy noise covariance by the secondary data. Then we construct a detector based on the estimated covariance, and further analyze the trade-off between privacy and detection probability under unknown privacy noise covariance.

Notations. Let $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{R}^n$ and $\mathbb{R}^{n\times m}$ be the sets of integer numbers, real numbers, n-dimensional real vectors and $n \times m$ real matrices, respectively. For any symmetric matrix P, the notation $P \succ 0$ ($P \succeq 0$) means that P is positive definite (semidefinite). The identity matrix is denoted as I with compatible dimension, respectively. For any matrix A, $\mathrm{Tr}(A)$ is used to denote the trace of A. The expectation of a random variable x is denoted by $\mathbb{E}(x)$. The notation $\mathcal{N}(\mu, \Sigma)$ represents a Gaussian distribution with mean value μ and covariance matrix Σ. Let χ^2_v and $\chi^2_v(c)$ be a central Chi-squared distribution and non-central Chi-squared distribution, respectively, where v is degree of freedom and c is non-centrality parameter. The notation $\mathrm{diag}_{j\in\mathcal{J}}[Q_j]$ is block diagonal concatenation matrices Q_j with j belonging to a set of indices $\mathcal{J}$. Let $\mathrm{col}_{j\in\mathcal{J}}[y_j]$ and $\mathrm{row}_{j\in\mathcal{J}}[y_j]$ be the column and row concatenation of vectors $y_j, j \in \mathcal{J}$, respectively. The same notation is also applied with matrices. For a sequence of vectors $y(i) \in \mathbb{R}^n$, $i = k_1, k_2, \ldots, k_s$, the vector $(y)_{k_1}^{k_s} = \mathrm{col}[y(k_1), y(k_2), \ldots, y(k_s)]$. For any $g \in \mathbb{Z}$, $(g)_k^+ = \prod_{i=0}^{k-1}(g+i)$, $k \geq 1$.

2. Preliminaries

In this section, the preliminaries related to system model, attack model and local filter are introduced.

2.1. System model

We consider a discrete-time interconnected system composed of N subsystems. Let $S \triangleq \{1, \ldots, N\}$ be the set of all subsystems and define the set of neighbors of subsystems i as $\mathcal{N}_i$. The dynamics of subsystem i are described as

$$x_i(k+1) = A_i x_i(k) + B_i u_i(k) + \sum_{j\in\mathcal{N}_i} A_{ij} x_j(k) + w_i(k), \tag{1}$$

$$y_i(k) = C_i x_i(k) + v_i(k), \tag{2}$$

where $x_i(k) \in \mathbb{R}^{n_i}$ is the state variable, $u_i(k) \in \mathbb{R}^{q_i}$ is the control input, and $y_i(k) \in \mathbb{R}^{m_i}$ is the sensor measurement of subsystem i. The process noise $w_i(k)$ and the measurement noise $v_i(k)$ are independent and identically distributed zero-mean Gaussian signals with covariance matrices $\Sigma_{w_i} \succeq 0$ and $\Sigma_{v_i} \succ 0$, respectively. The initial state $x_i(0)$ is a zero-mean Gaussian random variable with covariance $\Sigma_{x_i} \succ 0$, and is independent of $w_i(k)$ and $v_i(k)$. The matrices A_i, B_i, A_{ij} and C_i are real-valued with compatible dimensions. The pair (A_i, B_i) is controllable, and the pair (A_i, C_i) is observable.

Then following (Barboni et al., 2019), we can treat the interconnection term in (1) as unknown input. Set

$$\sum_{j\in\mathcal{N}_i} A_{ij} x_j(k) = E_i \zeta_i(k) = G_i \bar{E}_i \zeta_i(k) = G_i \xi_i(k), \tag{3}$$

where $E_i = \mathrm{row}_{j\in\mathcal{N}_i}[A_{ij}]$, $\zeta_i(k) = \mathrm{col}_{j\in\mathcal{N}_i}[x_j(k)]$, $\xi_i(k) = \bar{E}_i \zeta_i(k) \in \mathbb{R}^{g_i}$, G_i is a full column rank matrix and $\bar{E}_i$ is a weight matrix. Thus the dynamic (1) can be transformed into

$$x_i(k+1) = A_i x_i(k) + B_i u_i(k) + G_i \xi_i(k) + w_i(k). \tag{4}$$

2.2. Attack model

We consider the scenario that the attacker has the knowledge of the subsystem model (A_i, B_i, C_i), $\forall i \in S$, and has access to the original transmitted signals $u_i(k)$ and $y_i(k)$. Then the attacker modifies $u_i(k)$ and $y_i(k)$ into $\tilde{u}_i(k)$ and $\tilde{y}_i(k)$ by attack signals $\eta_i(k)$ and $\gamma_i(k)$, respectively, which is shown in Fig. 1. It follows from Smith (2015) that the attack signals $\eta_i(k)$ and $\gamma_i(k)$ can be modelled as

$$x_i^a(k+1) = A_i x_i^a(k) + B_i \eta_i(k), \tag{5}$$

$$\gamma_i(k) = C_i x_i^a(k), \tag{6}$$

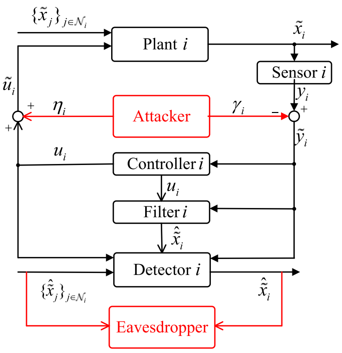


Fig. 1. Architecture of attacked subsystem i.

where $x_i^a(k)$ is the state of the attacker, $\eta_i(k)$ is an arbitrary signal injected to deteriorate the system performance, and $\gamma_i(k)$ is injected to eliminate the effect of the attack signal $\eta_i(k)$ on the measurement output. We assume that the attack signals begin at $k_a \geq 1$. Thus, it holds that $x_i^a(k) = 0$ for $k \leq k_a$. Denote $\tilde{x}_i(k)$ and $\tilde{y}_i(k)$ as the attacked state variable and sensor measurement of subsystem i, respectively. Then the dynamics of attacked subsystem i are described as

$$\tilde{x}_i(k+1) = A_i\tilde{x}_i(k) + B_i\tilde{u}_i(k) + \sum_{j\in\mathcal{N}_i} A_{ij}\tilde{x}_j(k) + w_i(k), \tag{7}$$

$$\tilde{y}_i(k) = C_i\tilde{x}_i(k) + v_i(k) - \gamma_i(k), \tag{8}$$

where $\tilde{u}_i(k) = u_i(k) + \eta_i(k)$. Moreover, it can be derived from (3), (4) and (7) that

$$\tilde{x}_i(k+1) = A_i\tilde{x}_i(k) + B_i\tilde{u}_i(k) + G_i\tilde{\xi}_i(k) + w_i(k), \tag{9}$$

where $\tilde{\xi}_i(k)$ satisfies $\sum_{j\in\mathcal{N}_i} A_{ij}\tilde{x}_j(k) = G_i\tilde{\xi}_i(k)$.

2.3. Local filter

We apply an unbiased minimum variance filter proposed in Gillijns and De Moor (2007) to estimate the state $x_i(k)$ and unknown term $\xi_i(k)$. Denote $\hat{x}_i(k)$ and $\hat{\xi}_i(k)$ as the estimations of $x_i(k)$ and $\xi_i(k)$, respectively. Then we have

$$\hat{x}_i(k) = \bar{A}_i[A_i\hat{x}_i(k-1) + B_iu_i(k-1)] + \bar{L}_iy_i(k), \tag{10}$$

and

$$\hat{\xi}_i(k-1) = M_i[y_i(k) - C_i(A_i\hat{x}_i(k-1) + B_iu_i(k-1))], \tag{11}$$

where $\bar{L}_i = K_i + (I - K_iC_i)G_iM_i$, and $\bar{A}_i = (I - K_iC_i)(I - G_iM_iC_i) = I - \bar{L}_iC_i$ with M_i and K_i gain matrices to be determined. Moreover, the following assumption is needed for the unbiased minimum variance filter.

Assumption 1 (*Gillijns & De Moor, 2007*). Each matrix C_i, $i \in S$, satisfies $\mathrm{Rank}(C_iG_i) = \mathrm{Rank}(G_i) = g_i$ with g_i the dimension of $\xi_i(k)$.

The estimation error of subsystem i under no attacks is defined as $e_i(k) = x_i(k) - \hat{x}_i(k)$. From (1) and (10), we have

$$e_i(k) = \hat{A}_ie_i(k-1) + \bar{A}_iw_i(k-1) - \bar{L}_iv_i(k), \tag{12}$$

where $\hat{A}_i = \bar{A}_iA_i$.

Then it can be derived that $e_i(k) \backsim \mathcal{N}(0, \Sigma_{e_i}(k))$ with $\Sigma_{e_i}(k) = \hat{A}_i\Sigma_{e_i}(k-1)\hat{A}_i^T + \bar{A}_i\Sigma_{w_i}\bar{A}_i^T + L_i\Sigma_{v_i}L_i^T$. The following lemma provides conditions for the stability of the filter.

Lemma 1 (*Fang & de Callafon, 2012*). *For each subsystem i, if there exist matrices M_i and K_i such that M_i satisfies $M_iC_iG_i = I$ and $|\lambda_j(\hat{A}_i)| < 1$, $j = 1, \ldots, n_i$, where $\lambda_j(\hat{A}_i)$ is j-th eigenvalue of $\hat{A}_i$, and if the pair $(A_i, \Sigma_{w_i}^{\frac{1}{2}})$ is controllable, then $\Sigma_{e_i}(k)$ converges to $\Sigma_{\bar{e}_i}$ for any initial $\Sigma_{e_i}(0)$, where $\Sigma_{\bar{e}_i}$ is the unique positive semi-definite solution of $\Sigma_{\bar{e}_i} = \hat{A}_i\Sigma_{\bar{e}_i}\hat{A}_i^T + \bar{A}_i\Sigma_{w_i}\bar{A}_i^T + L_i\Sigma_{v_i}L_i^T$.*

Without loss of generality, we assume that the filter starts from the steady state, i.e., $\Sigma_{e_i}(0) = \Sigma_{\bar{e}_i}$.

If the subsystem i is attacked, we have the attacked state estimation from (8) and (10) that

$$\hat{\tilde{x}}_i(k) = \bar{A}_i[A_i\hat{\tilde{x}}_i(k-1) + B_iu_i(k-1)] + \bar{L}_i\tilde{y}_i(k). \tag{13}$$

Then it follows from (5), (7) and (13) that the attacked estimation error is described as

$$\tilde{e}_i(k) = \tilde{x}_i(k) - \hat{\tilde{x}}_i(k) = e_i^r(k) + x_i^a(k), \tag{14}$$

where $e_i^r(k) = \hat{A}_ie_i^r(k-1) + \bar{A}_iw_i(k-1) - \bar{L}_iv_i(k)$. Therefore, the attacked estimation error $\tilde{e}_i(k)$ is affected by the attack signals. Moreover, we have $\tilde{e}_i(k) \backsim \mathcal{N}(x_i^a(k), \Sigma_{\bar{e}_i})$, $k \geq k_a$. Thus, it is necessary to construct the attack detector to detect the attack signals $\eta_i(k)$ and $\gamma_i(k)$.

3. Attack detector and privacy-preserving method

In this section, we firstly design a distributed attack detector to detect local covert attacks $\eta_i(k)$ and $\gamma_i(k)$. Moreover, as shown in Fig. 1, there exists an additional eavesdropper, which is able to infer the private state information, lurking within the communication between two neighboring subsystems. Therefore, the design of privacy-preserving method is also provided to protect private state information from the eavesdropper.

3.1. Distributed attack detector

From (7) and (8), we can observe that if at least one neighboring subsystem of subsystem i is attacked, the attacked state $\tilde{x}_j(k-1)$, $j \in \mathcal{N}_i$, can affect $\tilde{x}_i(k)$, and thus affect $\tilde{y}_i(k)$. If the state estimation $\hat{\tilde{x}}_j(k-1)$ of subsystem j is transmitted to subsystem i, then based on $\tilde{y}_i(k)$ and $\hat{\tilde{x}}_j(k-1)$, the subsystem i can apply a new residual $z_i(k)$ which will be affected by the attacked estimation error $\tilde{e}_j(k-1)$ for detection.

Then the residual $z_i(k)$ is constructed as

$$\begin{aligned} z_i(k) = y_i(k) - C_i[&A_i\hat{x}_i(k-1) + B_iu_i(k-1) \\ &+ \sum_{j\in\mathcal{N}_i} A_{ij}\hat{x}_j(k-1)]. \end{aligned} \tag{15}$$

The distribution of $z_i(k)$ is given in the following lemma.

Lemma 2. *The distribution of the residual $z_i(k)$ is described as*

$$z_i(k) \backsim \begin{cases} \mathcal{N}(0, \Sigma_{z_i}), & k < k_a, \\ \mathcal{N}(a_i(k-1), \Sigma_{z_i}), & k \geq k_a, \end{cases}$$

where $a_i(k-1) = C_iE_i col_{j\in\mathcal{N}_i}[x_j^a(k-1)]$, and $\Sigma_{z_i} = C_iA_i\Sigma_{\bar{e}_i}A_i^TC_i^T + C_iE_i diag_{j\in\mathcal{N}_i}[\Sigma_{\bar{e}_j}]E_i^TC_i^T + C_i\Sigma_{w_i}C_i^T + \Sigma_{v_i} \succ 0$.

Proof. The proof can be found in the arXiv version (Wang, Liu, Li, Fridman, & Xia, 2024).

Remark 1. From Lemma 2, we know that if at least one neighboring subsystem of subsystem i is attacked, then the expectation of $z_i(k)$ is affected by the attacks on subsystem j, $j \in \mathcal{N}_i$. Therefore, we can design an attack detector based on residual $z_i(k)$ to detect whether the neighboring subsystems are under attacks.

The attack detection problem can be described as a binary hypothesis testing problem. Let H_0 and H_1 be the hypothesis that the attacks are absent and present, respectively. Since subsystem i has no knowledge of vector $a_i(k-1)$, the Generalized Likelihood Ratio Test (GLRT) criterion is applied for the testing problem, which is described as

$$\frac{f(z_i(k)|H_0)}{\sup_{a_i(k-1)} f(z_i(k)|H_1)} \overset{H_0}{\underset{H_1}{\gtrless}} \tau_i', \tag{16}$$

where $f(z_i(k)|H_0)$ and $f(z_i(k)|H_1)$ are the probability density functions of $z_i(k)$ under hypotheses H_0 and H_1, respectively, and $\tau_i' > 0$ is a threshold. Following Kay (1993), we can transform (16) into

$$t(z_i(k)) \triangleq z_i^T(k)\Sigma_{z_i}^{-1}z_i(k) \overset{H_1}{\underset{H_0}{\gtrless}} \tau_i, \tag{17}$$

where $\tau_i > 0$ is the detection threshold needed to be determined.

Lemma 3. *It holds that $t(z_i(k)) \backsim \chi^2_{m_i}$ under H_0, and $t(z_i(k)) \backsim \chi^2_{m_i}(c_i(k))$ under H_1, where $c_i(k) = a_i^T(k-1)\Sigma_{z_i}^{-1}a_i(k-1)$ is a non-centrality parameter.*

Proof. The proof can be found in the arXiv version (Wang et al., 2024).

We adopt the false alarm probability and the detection probability to describe the detection performance of detector (17), which are given by

$$P_{i,f} = P(t(z_i(k)) > \tau_i|H_0) = 1 - F_{m_i}(\tau_i), \tag{18}$$

and

$$P_{i,d}(k) = P(t(z_i(k)) > \tau_i|H_1) = 1 - F_{m_i}(\tau_i, c_i(k)), \tag{19}$$

respectively, where $F_{m_i}(\tau_i)$ is the Cumulative Distribution Function (CDF) of central Chi-squared distribution $\chi^2_{m_i}$ and $F_{m_i}(\tau_i, c_i(k))$ is the CDF of non-central Chi-squared distribution $\chi^2_{m_i}(c_i(k))$.

3.2. Privacy concern

In order to protect the privacy of subsystem j when transmitting state estimation to subsystem i, we design the privacy-preserving method as follows

$$\theta_j^i(k) = \hat{\tilde{x}}_j(k) + \alpha_j^i(k), \tag{20}$$

where $\theta_j^i(k)$ is the noisy state estimation, $\alpha_j^i(k) \backsim \mathcal{N}(0, \Sigma_{\alpha_j^i})$ is the privacy noise, and the covariance $\Sigma_{\alpha_j^i} \succ 0$ needs to be designed.

To quantify the privacy, we use the mutual information $I[(\tilde{x}_j)_1^K; (\theta_j^i)_1^K]$ between private and disclosed information from instants 1 to K. Then, if we only focus on privacy preservation performance, the optimal noise covariance can be obtained by solving the optimization problem

$$\begin{aligned} &\min_{\{\Sigma_{\alpha_j^i}\}_{j\in\mathcal{N}_i}} \sum_{j\in\mathcal{N}_i} I[(\tilde{x}_j)_1^K; (\theta_j^i)_1^K] \\ &\quad s.t. \ \ \Sigma_{\alpha_j^i} \succ 0, j \in \mathcal{N}_i. \end{aligned} \tag{21}$$

Therefore, it is necessary to formulate the mutual information $I[(\tilde{x}_j)_1^K; (\theta_j^i)_1^K], j \in \mathcal{N}_i$, in terms of the privacy noise covariance $\Sigma_{\alpha_j^i}$. Following Cover and Thomas (1991), we can describe the mutual information $I[(\tilde{x}_j)_1^K; (\theta_j^i)_1^K]$ as

$$I[(\tilde{x}_j)_1^K; (\theta_j^i)_1^K] = H[(\tilde{x}_j)_1^K] + H[(\theta_j^i)_1^K] - H[(\tilde{x}_j)_1^K, (\theta_j^i)_1^K], \tag{22}$$

where $H[(\tilde{x}_j)_1^K]$ and $H[(\theta_j^i)_1^K]$ are differential entropy of $(\tilde{x}_j)_1^K$ and $(\theta_j^i)_1^K$, respectively, and $H[(\tilde{x}_j)_1^K, (\theta_j^i)_1^K]$ is joint entropy. For the convenience of analysis, we define

$$\Psi(\Xi) = \begin{bmatrix} \Xi & 0 & \dots & 0 \\ A_j\Xi & \Xi & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A_j^{K-1}\Xi & A_j^{K-2}\Xi & \dots & \Xi \end{bmatrix}$$

and

$$\hat{\Psi}(\Xi) = \begin{bmatrix} \Xi & 0 & \dots & 0 \\ \hat{A}_j\Xi & \Xi & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{A}_j^{K-1}\Xi & \hat{A}_j^{K-2}\Xi & \dots & \Xi \end{bmatrix},$$

where $\hat{A}_j = \bar{A}_jA_j$. Then let $\Psi_{\tilde{u}_j} = \Psi(B_j)$, $\Psi_{\tilde{\xi}_j} = \Psi(G_j)$, $\Psi_{w_j} = \Psi(I)$, $\hat{\Psi}_{w_j} = \hat{\Psi}(\bar{A}_j)$ and $\hat{\Psi}_{v_j} = \hat{\Psi}(L_j)$.

Lemma 4. *It holds that*

$$\begin{bmatrix} (\theta_j^i)_1^K \\ (\tilde{x}_j)_1^K \end{bmatrix} \backsim \mathcal{N}\left(\begin{bmatrix} \mu_{\theta_j^i} \\ \mu_{\tilde{x}_j} \end{bmatrix}, \begin{bmatrix} \Sigma_{\theta_j^i} & \Sigma_{\tilde{x}_j\theta_j^i}^T \\ \Sigma_{\tilde{x}_j\theta_j^i} & \Sigma_{\tilde{x}_j} \end{bmatrix}\right),$$

where

$$\begin{aligned} \mu_{\tilde{x}_j} &= \Psi_{\tilde{u}_j}(\tilde{u}_j)_0^{K-1} + \Psi_{\tilde{\xi}_j}(\tilde{\xi}_j)_0^{K-1}, \\ \mu_{\theta_j^i} &= \Psi_{\tilde{u}_j}(\tilde{u}_j)_0^{K-1} + \Psi_{\tilde{\xi}_j}(\tilde{\xi}_j)_0^{K-1} - (x_j^a)_1^K, \\ \Sigma_{\tilde{x}_j} &= \Theta_j\Sigma_{x_j}\Theta_j^T + \Psi_{w_j}\check{\Sigma}_{w_j}\Psi_{w_j}^T, \qquad (23) \\ \Sigma_{\theta_j^i} &= \Theta_j\Sigma_{x_j}\Theta_j^T + \hat{\Theta}_j\Sigma_{e_j}\hat{\Theta}_j^T - \Theta_j\Sigma_{x_j}\hat{\Theta}_j^T \\ &\quad - \hat{\Theta}_j\Sigma_{x_j}\Theta_j^T + (\Psi_{w_j} - \hat{\Psi}_{w_j})\check{\Sigma}_{w_j}(\Psi_{w_j} - \hat{\Psi}_{w_j})^T \\ &\quad + \hat{\Psi}_{v_j}\check{\Sigma}_{v_j}\hat{\Psi}_{v_j}^T + \check{\Sigma}_{\alpha_j^i}, \qquad (24) \end{aligned}$$

and $\Sigma_{\tilde{x}_j\theta_j^i} = \Theta_j\Sigma_{x_j}\Theta_j^T - \Theta_j\Sigma_{x_j}\hat{\Theta}_j^T + \Psi_{w_j}\check{\Sigma}_{w_j}(\Psi_{w_j} - \hat{\Psi}_{w_j})^T$ with $\check{\Sigma}_{w_j} = I \otimes \Sigma_{w_j}$, $\check{\Sigma}_{v_j} = I \otimes \Sigma_{v_j}$ $\check{\Sigma}_{\alpha_j^i} = I \otimes \Sigma_{\alpha_j^i}$, $\Theta_j = col[A_j, A_j^2, \dots, A_j^K]$, and $\hat{\Theta}_j = col[\hat{A}_j, \hat{A}_j^2, \dots, \hat{A}_j^K]$.

Proof. The proof can be found in the arXiv version (Wang et al., 2024).

From Lemma 4 and (22), we get

$$\begin{aligned} I[(\tilde{x}_j)_1^K; (\theta_j^i)_1^K] = \frac{1}{2}\big[&-\log\det(\Sigma_{\tilde{x}_j} - \Sigma_{\tilde{x}_j\theta_j^i}\Sigma_{\theta_j^i}^{-1}\Sigma_{\tilde{x}_j\theta_j^i}^T) \\ &+ \log\det(\Sigma_{\tilde{x}_j})\big], \end{aligned} \tag{25}$$

where $\Sigma_{\theta_j^i}$ contains the privacy noise covariance $\Sigma_{\alpha_j^i}$. Therefore, the mutual information has been formulated in terms of the privacy noise distribution. Then, by the monotonicity of determinant, the optimization problem (21) can be rewritten as

$$\begin{aligned} &\min_{\{\Gamma_j, \Sigma_{\alpha_j^i}\}_{j\in\mathcal{N}_i}} \sum_{j\in\mathcal{N}_i} -\log\det(\Gamma_j) \\ &\quad s.t. \ \begin{cases} \Sigma_{\tilde{x}_j} - \Sigma_{\tilde{x}_j\theta_j^i}\Sigma_{\theta_j^i}^{-1}\Sigma_{\tilde{x}_j\theta_j^i}^T \succeq \Gamma_j \succ 0, \\ \Sigma_{\alpha_j^i} \succ 0, j \in \mathcal{N}_i. \end{cases} \end{aligned} \tag{26}$$

Furthermore, by Schur complement, (26) is equivalent to the following convex optimization problem

$$\min_{\{\Gamma_j, \Sigma_{\alpha_j^i}\}_{j\in\mathcal{N}_i}} \sum_{j\in\mathcal{N}_i} -\log\det(\Gamma_j) \tag{27}$$

$$s.t.\quad \begin{cases} \begin{bmatrix} \Sigma_{\tilde{x}_j} - \Gamma_j & \Sigma_{\tilde{x}_j\theta_j^i} \\ \Sigma_{\tilde{x}_j\theta_j^i}^T & \Sigma_{\theta_j^i} \end{bmatrix} \succeq 0, \\ \Gamma_j \succ 0,\ \Sigma_{\alpha_j^i} \succ 0, j \in \mathcal{N}_i. \end{cases}$$

Moreover, after adding the privacy noise, the residual of subsystem i under attacks is given by

$$\begin{aligned} z_i^p(k) = \tilde{y}_i(k) - C_i[A_i\hat{\tilde{x}}_i(k-1) + B_iu_i(k-1) \\ + \sum_{j\in\mathcal{N}_i} A_{ij}(\hat{\tilde{x}}_j(k-1) + \alpha_j^i(k-1))]. \end{aligned} \tag{28}$$

From Lemma 2 and (20), we have $z_i^p(k) \backsim \mathcal{N}(C_iE_ia_i(k-1), \Sigma_{z_i^p})$, where

$$\Sigma_{z_i^p} = \Sigma_{z_i} + \Sigma_{p_i} \tag{29}$$

with $\Sigma_{p_i} = C_iE_i\text{diag}_{j\in\mathcal{N}_i}[\Sigma_{\alpha_j^i}]E_i^TC_i^T$. Then the detector (17) is transformed into

$$t(z_i^p(k)) \triangleq [z_i^p(k)]^T \Sigma_{z_i^p}^{-1} z_i^p(k) \underset{H_0}{\overset{H_1}{\gtrless}} \tau_i. \tag{30}$$

Therefore, the privacy-preserving method can affect the distribution of the residual $z_i^p(k)$, thereby affecting the CDF of $t(z_i^p(k))$. Thus the detection performance such as detection probability and the false alarm probability of subsystem i may be affected by the privacy noise $\alpha_j^i(k), j \in \mathcal{N}_i$.

Furthermore, to increase the degree of privacy preservation, the privacy noise covariance $\Sigma_{\alpha_j^i}$ from neighboring subsystems may be unknown to subsystem i. Therefore, subsystem i may still use the covariance Σ_{z_i} to construct the detector, which is given by

$$t^p(z_i^p(k)) \triangleq [z_i^p(k)]^T \Sigma_{z_i}^{-1} z_i^p(k) \underset{H_0}{\overset{H_1}{\gtrless}} \tau_i. \tag{31}$$

It can be seen from (30) and (31) that the detection variables $t(z_i^p(k))$ and $t^p(z_i^p(k))$ are different, which means that the CDFs corresponding to $t(z_i^p(k))$ and $t^p(z_i^p(k))$ are also different. Therefore, the detectors (30) and (31) may have different detection performance. It is necessary to analyze the trade-off between privacy and security under known and unknown privacy noise, respectively.

4. The trade-off between privacy and security under known privacy noise covariance

In this section, we firstly analyze the effects of privacy noise on the false alarm probability and the detection probability for the detector (30) with known privacy noise covariance. Moreover, on the basis of the optimization problem (27), in order to increase the detection performance, we reformulate an optimization problem to obtain the covariance of the privacy noise.

We firstly give the following lemma to describe the distribution of detection variable $t(z_i^p(k))$ given in (30).

Lemma 5. *It holds that $t(z_i^p(k)) \backsim \chi_{m_i}^2$ under H_0, and $t(z_i^p(k)) \backsim \chi_{m_i}^2(c_i^p(k))$ under H_1, where $\Sigma_{z_i^p}$ is given in (29) and*

$$c_i^p(k) = a_i^T(k-1)\Sigma_{z_i^p}^{-1}a_i(k-1) \tag{32}$$

is non-centrality parameter.

Proof. The proof can be found in the arXiv version (Wang et al., 2024).

Remark 2. From Lemmas 3 and 5, we can obtain that under H_0, the detection variables $t(z_i^p(k))$ and $t(z_i(k))$ follow the same central Chi-squared distribution $\chi_{m_i}^2$. Therefore, if each subsystem shares the covariance of the privacy noise with its neighboring subsystems, the false alarm probability will not increase.

Inspired by Neyman-Pearson test criterion (Kay, 1993), we need to preset false-alarm rate threshold p_i^f and to determine the detection threshold τ_i. Then, it follows from Katewa et al. (2021) that

$$\tau_i = 2P_g^{-1}(\frac{m_i}{2}, 1 - p_i^f), \tag{33}$$

where $P_g^{-1}(\cdot,\cdot)$ is the inverse regularized lower incomplete Gamma function. From Lemma 5 and (19), we have the detection probability with the privacy noise as follows

$$P_{i,d}^p(k) = P(t(z_i^p(k)) > \tau_i | H_1) = 1 - F_{m_i}(\tau_i, c_i^p(k)). \tag{34}$$

Therefore, the detection probability $P_{i,d}^p(k)$ is dependent on non-centrality parameter $c_i^p(k)$. Furthermore, it can be observed from (29) and (32) that $c_i^p(k)$ is affected by $\text{diag}_{j\in\mathcal{N}_i}[\Sigma_{\alpha_j^i}]$. Thus, the privacy noise α_j^i from neighboring subsystems of subsystem i can affect the detection probability $P_{i,d}^p(k)$.

Then we give the following theorem to describe the trade-off between the privacy and the detection probability under known privacy noise covariance.

Theorem 1. *If the privacy noise covariances $\Sigma_{\alpha_j^i}^{(a)}$ and $\Sigma_{\alpha_j^i}^{(b)}, j \in \mathcal{N}_i$, satisfy $\Sigma_{\alpha_j^i}^{(a)} \succeq \Sigma_{\alpha_j^i}^{(b)} \succ 0$, then we have $I^{(a)}[(\tilde{x}_j)_1^K; (\theta_j^i)_1^K] \leq I^{(b)}[(\tilde{x}_j)_1^K; (\theta_j^i)_1^K]$ while $P_{i,d}^{p(a)}(k) \leq P_{i,d}^{p(b)}(k)$.*

Proof. Due to $\Sigma_{\alpha_j^i}^{(a)} \succeq \Sigma_{\alpha_j^i}^{(b)}$, then by (24), we have $\Sigma_{\theta_j^i}^{(a)} - \Sigma_{\theta_j^i}^{(b)} \succeq 0$, which means that $(\Sigma_{\theta_j^i}^{(b)})^{-1} - (\Sigma_{\theta_j^i}^{(a)})^{-1} \succeq 0$ (Horn & Johnson, 2012). Thus, it follows that $\Sigma_{\tilde{x}_j\theta_j^i}[(\Sigma_{\theta_j^i}^{(b)})^{-1} - (\Sigma_{\theta_j^i}^{(a)})^{-1}]\Sigma_{\tilde{x}_j\theta_j^i}^T \succeq 0$. Then we derive that $\Sigma_{\tilde{x}_j} - \Sigma_{\tilde{x}_j\theta_j^i}(\Sigma_{\theta_j^i}^{(a)})^{-1}\Sigma_{\tilde{x}_j\theta_j^i}^T \succeq \Sigma_{\tilde{x}_j} - \Sigma_{\tilde{x}_j\theta_j^i}(\Sigma_{\theta_j^i}^{(b)})^{-1}\Sigma_{\tilde{x}_j\theta_j^i}^T$. Due to the monotonicity of determinant, it can be obtained that $\det(\Sigma_{\tilde{x}_j} - \Sigma_{\tilde{x}_j\theta_j^i}(\Sigma_{\theta_j^i}^{(a)})^{-1}\Sigma_{\tilde{x}_j\theta_j^i}^T) \geq \det(\Sigma_{\tilde{x}_j} - \Sigma_{\tilde{x}_j\theta_j^i}(\Sigma_{\theta_j^i}^{(b)})^{-1}\Sigma_{\tilde{x}_j\theta_j^i}^T)$. From (25), we have $I^{(a)}[(\tilde{x}_j)_1^K; (\theta_j^i)_1^K] \leq I^{(b)}[(\tilde{x}_j)_1^K; (\theta_j^i)_1^K]$.

Following Ghosh (1973), we can describe $F_{m_i}(\tau_i, c_i^p(k))$ in (34) as $F_{m_i}(\tau_i, c_i^p(k)) = e^{-c_i^p(k)/2}\sum_{t=0}^{\infty}\left[\frac{(c_i^p(k)/2)^t}{t!}F_{m_i+2t}(\tau_i)\right]$, where $F_{m_i+2t}(\tau_i)$ is the CDF of $\chi_{m_i+2t}^2$ with $m_i + 2t$ degrees of freedom. Since $F_{m_i}(\tau_i, c_i^p(k))$ is a decreasing function of non-centrality parameter $c_i^p(k)$ (Katewa et al., 2021), the detection probability $P_{i,d}^p(k)$ is an increasing function of $c_i^p(k)$.

Due to $\Sigma_{\alpha_j^i}^{(a)} \succeq \Sigma_{\alpha_j^i}^{(b)}, j \in \mathcal{N}_i$, then $C_iE_i\text{diag}_{j\in\mathcal{N}_i}[\Sigma_{\alpha_j^i}^{(a)} - \Sigma_{\alpha_j^i}^{(b)}]E_i^TC_i^T \succeq 0$. It follows from (29) that $\Sigma_{z_i^p}^{(a)} \succeq \Sigma_{z_i^p}^{(b)}$, which results in

$$\begin{aligned} [c_i^p(k)]^{(a)} &= a_i^T(k-1)(\Sigma_{z_i^p}^{(a)})^{-1}a_i(k-1) \\ &\leq a_i^T(k-1)(\Sigma_{z_i^p}^{(b)})^{-1}a_i(k-1) \\ &= [c_i^p(k)]^{(b)}. \end{aligned} \tag{35}$$

Therefore, we have $P_{i,d}^{p(a)}(k) \leq P_{i,d}^{p(b)}(k)$. ■

Remark 3. Theorem 1 shows that if neighboring subsystems increase the privacy noise covariance, the mutual information will decrease. Thus, the degree of privacy preservation will increase. However, the detection performance will decrease. Therefore, there is a trade-off between privacy and security.

In order to increase the detection probability, it is necessary to increase the value of $c_i^p(k)$. Intuitively, $c_i^p(k)$ is larger if $\mathrm{Tr}(\Sigma_{z_i^p}^{-1})$ is larger. Moreover, maximizing $\mathrm{Tr}(\Sigma_{z_i^p}^{-1})$ can be achieved by minimizing $\mathrm{Tr}(\Sigma_{z_i^p})$. Eq. (29) means that minimizing $\mathrm{Tr}(\Sigma_{z_i^p})$ is equivalent to minimize $\mathrm{Tr}(\Sigma_{p_i})$. Therefore, from (27) we give the following optimization problem to obtain the privacy noise covariance

$$\min_{\{\Gamma_j,\Sigma_{\alpha_j^i}\}_{j\in\mathcal{N}_i}} \sum_{j\in\mathcal{N}_i} -\log\det(\Gamma_j)+\kappa_i\mathrm{Tr}(\Sigma_{p_i}) \tag{36}$$

$$s.t. \quad \begin{cases} \Sigma_{p_i}=C_iE_i\mathrm{diag}_{j\in\mathcal{N}_i}[\Sigma_{\alpha_j^i}]E_i^TC_i^T, \\ \begin{bmatrix} \Sigma_{\tilde{x}_j}-\Gamma_j & \Sigma_{\tilde{x}_j\theta_j^i} \\ \Sigma_{\tilde{x}_j\theta_j^i}^T & \Sigma_{\theta_j^i} \end{bmatrix} \succeq 0, \\ \Gamma_j \succ 0,\ \Sigma_{\alpha_j^i}\succ 0, j\in\mathcal{N}_i, \end{cases}$$

where $\kappa_i > 0$ is a weight factor.

5. The trade-off between privacy and security under unknown privacy noise covariance

In this section, we firstly analyze the effects of privacy noise on the false alarm probability and the detection probability for the detector (31) with unknown privacy noise covariance, respectively. Furthermore, an optimization problem with guaranteeing the detection performance is established to obtain the privacy noise covariance.

5.1. False alarm probability under unknown privacy noise covariance

In (28), the residual $z_i^p(k)$ follows $\mathcal{N}(0,\Sigma_{z_i^p})$ under H_0, where $\Sigma_{z_i^p}$ is given in (29). Thus, the detection variable $t^p(z_i^p(k))$ in (31) no longer follows the central Chi-squared distribution. Therefore, the false alarm probability can be affected by the privacy noise.

Lemma 6. *If subsystem i has no knowledge of the privacy noise covariance $\Sigma_{\alpha_j^i}, j\in\mathcal{N}_i$, then we have $P_{i,f}^u \geq P_{i,f}$, where $P_{i,f}$ is given in (18), and $P_{i,f}^u$ is false alarm probability given by $P_{i,f}^u = P(t^p(z_i^p(k)) > \tau_i|H_0)$. Moreover, if C_iE_i is full row rank, then $P_{i,f}^u > P_{i,f}$.*

Proof. The proof can be found in the arXiv version (Wang et al., 2024).

Theorem 2. *If the matrices Σ_{p_i} and $\Sigma_{z_i}^{-1}$ are commutative and the privacy noise covariances $\Sigma_{\alpha_j^i}^{(a)}$ and $\Sigma_{\alpha_j^i}^{(b)}$, $j\in\mathcal{N}_i$, satisfy $\Sigma_{\alpha_j^i}^{(a)} \succeq \Sigma_{\alpha_j^i}^{(b)} \succ 0$, then we have $P_{i,f}^{u(a)} \geq P_{i,f}^{u(b)}$.*

Proof. Since Σ_{p_i} and $\Sigma_{z_i}^{-1}$ are commutative, then it holds that $\Sigma_{z_i}^{-1}\Sigma_{p_i} = \Sigma_{p_i}\Sigma_{z_i}^{-1}$, which means that $(\Sigma_{z_i}+\Sigma_{p_i})\Sigma_{z_i}^{-1} = \Sigma_{z_i}^{-1}(\Sigma_{p_i}+\Sigma_{z_i})$. Thus, $\Sigma_{z_i}^{-1}$ and $\Sigma_{p_i}+\Sigma_{z_i}$ are commutative. Moreover, $\Sigma_{z_i}^{-1}$ and $\Sigma_{p_i}+\Sigma_{z_i}$ are symmetric matrices. Therefore, $\Sigma_{z_i}^{-1}$ and $\Sigma_{p_i}+\Sigma_{z_i}$ have the same eigenvectors $\upsilon_1,\ldots,\upsilon_{m_i}$, i.e., $\Sigma_{z_i}^{-1} = \Omega_i\Phi_i\Omega_i^{-1}$ and $\Sigma_{p_i}+\Sigma_{z_i} = \Omega_i\Lambda_i\Omega_i^{-1}$, where $\Omega_i = [\upsilon_1,\ldots,\upsilon_{m_i}]$ is a matrix composed of eigenvectors, Φ_i and Λ_i are diagonal matrices with eigenvalues in the diagonal of $\Sigma_{z_i}^{-1}$ and $\Sigma_{p_i}+\Sigma_{z_i}$, respectively. Then we have $\Sigma_{z_i}^{-\frac{1}{2}} = \Omega_i\Phi_i^{\frac{1}{2}}\Omega_i^{-1}$ and $(\Sigma_{p_i}+\Sigma_{z_i})^{\frac{1}{2}} = \Omega_i\Lambda_i^{\frac{1}{2}}\Omega_i^{-1}$. It can be derived that $\Sigma_{z_i}^{-\frac{1}{2}}(\Sigma_{p_i}+\Sigma_{z_i})^{\frac{1}{2}} = \Omega_i\Phi_i^{\frac{1}{2}}\Lambda_i^{\frac{1}{2}}\Omega_i^{-1} = [\Sigma_{z_i}^{-1}(\Sigma_{p_i}+\Sigma_{z_i})]^{\frac{1}{2}}$ and $(\Sigma_{p_i}+\Sigma_{z_i})^{\frac{1}{2}}\Sigma_{z_i}^{-\frac{1}{2}} = \Omega_i\Lambda_i^{\frac{1}{2}}\Phi_i^{\frac{1}{2}}\Omega_i^{-1} = [(\Sigma_{p_i}+\Sigma_{z_i})\Sigma_{z_i}^{-1}]^{\frac{1}{2}}$. Therefore, we have

$$\begin{aligned} \Sigma_{z_i^p}^{\frac{1}{2}}\Sigma_{z_i}^{-1}\Sigma_{z_i^p}^{\frac{1}{2}} &= [(\Sigma_{p_i}+\Sigma_{z_i})^{\frac{1}{2}}\Sigma_{z_i}^{-\frac{1}{2}}][\Sigma_{z_i}^{-\frac{1}{2}}(\Sigma_{p_i}+\Sigma_{z_i})^{\frac{1}{2}}] \\ &= [(\Sigma_{p_i}+\Sigma_{z_i})\Sigma_{z_i}^{-1}]^{\frac{1}{2}}[\Sigma_{z_i}^{-1}(\Sigma_{p_i}+\Sigma_{z_i})]^{\frac{1}{2}} \\ &= (\Sigma_{p_i}+\Sigma_{z_i})\Sigma_{z_i}^{-1} \\ &= I+\Sigma_{p_i}\Sigma_{z_i}^{-1}. \end{aligned} \tag{37}$$

Define a Gaussian vector $\varphi_i(k) \backsim \mathcal{N}(0,I)$. Then the residual $z_i^p(k)$ in (28) can be described as $z_i^p(k) = \Sigma_{z_i^p}^{\frac{1}{2}}\varphi_i(k)$. From (31), we obtain

$$t^p(z_i^p(k)) = \varphi_i^T(k)\Sigma_{z_i^p}^{\frac{1}{2}}\Sigma_{z_i}^{-1}\Sigma_{z_i^p}^{\frac{1}{2}}\varphi_i(k). \tag{38}$$

Thus under the privacy noise covariance $\Sigma_{\alpha_j^i}^{(a)}$ and $\Sigma_{\alpha_j^i}^{(b)}$, the detection variables $[t^p(z_i^p(k))]^{(a)}$ and $[t^p(z_i^p(k))]^{(b)}$ can be given by

$$[t^p(z_i^p(k))]^{(a)} = \varphi_i^T(k)[\Sigma_{z_i^p}^{(a)}]^{\frac{1}{2}}\Sigma_{z_i}^{-1}[\Sigma_{z_i^p}^{(a)}]^{\frac{1}{2}}\varphi_i(k),$$

and

$$[t^p(z_i^p(k))]^{(b)} = \varphi_i^T(k)[\Sigma_{z_i^p}^{(b)}]^{\frac{1}{2}}\Sigma_{z_i}^{-1}[\Sigma_{z_i^p}^{(b)}]^{\frac{1}{2}}\varphi_i(k),$$

respectively, where $\Sigma_{z_i^p}^{(\ell)} = \Sigma_{z_i}+\Sigma_{p_i}^{(\ell)}$ with $\Sigma_{p_i}^{(\ell)} = C_iE_i\mathrm{diag}_{j\in\mathcal{N}_i}[\Sigma_{\alpha_j^i}^{(\ell)}]E_i^TC_i^T$, $\ell\in\{a,b\}$. By (37), we get $[t^p(z_i^p(k))]^{(a)} - [t^p(z_i^p(k))]^{(b)} = \varphi_i^T(k)(\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)})\Sigma_{z_i}^{-1}\varphi_i(k)$.

Because of $\Sigma_{\alpha_j^i}^{(a)} - \Sigma_{\alpha_j^i}^{(b)} \succeq 0$, then it holds that $\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)} = C_iE_i\mathrm{diag}_{j\in\mathcal{N}_i}[\Sigma_{\alpha_j^i}^{(a)} - \Sigma_{\alpha_j^i}^{(b)}]E_i^TC_i^T \succeq 0$. Due to $(\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)})\Sigma_{z_i}^{-1} = \Sigma_{z_i}^{\frac{1}{2}}[\Sigma_{z_i}^{-\frac{1}{2}}(\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)})\Sigma_{z_i}^{-\frac{1}{2}}]\Sigma_{z_i}^{-\frac{1}{2}}$, then $(\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)})\Sigma_{z_i}^{-1}$ and $\Sigma_{z_i}^{-\frac{1}{2}}(\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)})\Sigma_{z_i}^{-\frac{1}{2}}$ are similar. Therefore, all the eigenvalues of $(\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)})\Sigma_{z_i}^{-1}$ are not less than zero. Moreover, because Σ_{p_i} and $\Sigma_{z_i}^{-1}$ are commutative, then we have $[(\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)})\Sigma_{z_i}^{-1}]^T = [\Sigma_{z_i}^{-1}(\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)})]^T = (\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)})^T[\Sigma_{z_i}^{-1}]^T = (\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)})\Sigma_{z_i}^{-1}$. Therefore, $(\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)})\Sigma_{z_i}^{-1}$ is a symmetric matrix. Thus, we have $(\Sigma_{p_i}^{(a)}-\Sigma_{p_i}^{(b)})\Sigma_{z_i}^{-1} \succeq 0$, which means $[t^p(z_i^p(k))]^{(a)} - [t^p(z_i^p(k))]^{(b)} \geq 0$. The proof is completed. ■

Remark 4. Theorem 2 means that if the degree of privacy preservation is higher, the false alarm probability is larger under the condition that Σ_{p_i} and $\Sigma_{z_j}^{-1}$ are commutative. It is noted that this condition is only sufficient but not necessary. Therefore, even if the condition is not satisfied, the monotonically increasing relationship between the false alarm probability and the degree of privacy preservation may still hold.

From Lemma 6 and Theorem 2, we can conclude that adding the privacy noise can affect the false alarm probability. Therefore, to constrain the effects of privacy noise on the false alarm probability, we set the upper bound of false alarm distortion level by ν_i for subsystem $i\in S$, i.e., $P_{i,f}^u \leq p_i^f + \nu_i$, where p_i^f is given in (33). Then we have

$$F_{t^p(z_i^p(k))}(\tau_i) > 1 - p_i^f - \nu_i, \tag{39}$$

where $F_{t^p(z_i^p(k))}(\tau_i)$ is CDF of $t^p(z_i^p(k))$. However, according to (31), $t^p(z_i^p(k))$ does not follow the Chi-square distribution and there is no closed-form expression of its CDF. Our solution is to find the lower bound of $F_{t^p(z_i^p(k))}(\tau_i)$ and let the lower bound be greater than $1 - p_i^f - \nu_i$. Then we define a vector $\varrho_i(k) = h_i \varphi_i^T(k)\varphi_i(k)$, where h_i needs to be determined and $\varphi_i(k) \backsim \mathcal{N}(0, I)$. By (38), $t^p(z_i^p(k)) \le \varrho_i(k)$ if and only if $\Sigma_{z_i^p}^{\frac{1}{2}} \Sigma_{z_i}^{-1} \Sigma_{z_i^p}^{\frac{1}{2}} \preceq h_i I$. Following Hayati et al. (2024), we observe that if $t^p(z_i^p(k)) \le \varrho_i(k)$, then $F_{t^p(z_i^p(k))}(\tau_i) \ge F_{\varrho_i(k)}(\tau_i)$, where $F_{\varrho_i(k)}(\tau_i)$ is CDF of $\varrho_i(k)$. Therefore, if

$$\Sigma_{z_i^p}^{\frac{1}{2}} \Sigma_{z_i}^{-1} \Sigma_{z_i^p}^{\frac{1}{2}} \preceq h_i I \text{ and}$$
$$F_{\varrho_i(k)}(\tau_i) > 1 - p_i^f - \nu_i, \tag{40}$$

then condition (39) holds. Due to $\varphi_i^T(k)\varphi_i(k) \backsim \chi_{m_i}^2$, then we obtain $F_{\varrho_i(k)}(\tau_i) = P_g(\frac{m_i}{2}, \frac{\tau_i}{2h_i})$ with $P_g(\frac{m_i}{2}, \frac{\tau_i}{2h_i})$ being regularized Gamma function which is an increasing function of $\frac{\tau_i}{2h_i}$. Thus in order to satisfy (40), we can set $h_i < h_i^* = \frac{\tau_i}{2P_g^{-1}(\frac{m_i}{2}, 1-p_i^f-\nu_i)}$, where $P_g^{-1}(\cdot,\cdot)$ is the inverse of the lower incomplete Gamma function. Therefore, if we let $\Sigma_{z_i^p}^{\frac{1}{2}} \Sigma_{z_i}^{-1} \Sigma_{z_i^p}^{\frac{1}{2}} \preceq h_i I \prec h_i^* I$, then condition (39) holds. Moreover, $\Sigma_{z_i^p}^{\frac{1}{2}} \Sigma_{z_i}^{-1} \Sigma_{z_i^p}^{\frac{1}{2}} \prec h_i^* I$ is equivalent to

$$\Sigma_{z_i^p} \prec h_i^* \Sigma_{z_i}. \tag{41}$$

Therefore, integrating (27), (29) and (41), we can give the following optimization problem to obtain the private noise covariance

$$\min_{\{\Gamma_j, \Sigma_{\alpha_j^i}\}_{j\in\mathcal{N}_i}} \sum_{j\in\mathcal{N}_i} -\log\det(\Gamma_j) \tag{42}$$
$$s.t. \begin{cases} \begin{bmatrix} \Sigma_{\tilde{x}_j} - \Gamma_j & \Sigma_{\tilde{x}_j\theta_j^i} \\ \Sigma_{\tilde{x}_j\theta_j^i}^T & \Sigma_{\theta_j^i} \end{bmatrix} \succeq 0, \\ \Sigma_{z_i} + \Sigma_{p_i} \prec h_i^* \Sigma_{z_i}, \\ \Sigma_{p_i} = C_i E_i \mathrm{diag}_{j\in\mathcal{N}_i}[\Sigma_{\alpha_j^i}] E_i^T C_i^T, \\ h_i^* = \frac{\tau_i}{2P_g^{-1}(\frac{m_i}{2}, 1-p_i^f-\nu_i)}, \\ \Gamma_j \succ 0,\ \Sigma_{\alpha_j^i} \succ 0, j \in \mathcal{N}_i. \end{cases}$$

If ν_i is larger, then the false alarm probability is allowed to increase to a higher degree.

Since there is no closed-form expression of $F_{t^p(z_i^p(k))}(\tau_i)$, it is difficult to obtain the optimal covariance of the privacy noise. Alternatively, we can find a suboptimal covariance by solving the optimization problem (42) with the relaxation of the constraint (39). Moreover, the mutual information corresponding to the suboptimal covariance is an upper bound on that corresponding to the optimal covariance.

Remark 5. In Hayati et al. (2024), in order to analyze the trade-off between privacy and security, an optimization problem with maximizing privacy preservation performance while guaranteeing a bound on the false alarm probability is established to obtain the privacy noise covariance. In our paper, we not only establish the optimization problem (42) to obtain the privacy noise covariance, but also provide the theoretical analysis on the trade-off between privacy and security in Lemma 6 and Theorem 2.

5.2. Detection probability under unknown privacy noise covariance

In (28), the residual $z_i^p(k)$ follows $\mathcal{N}(C_i E_i a_i(k-1), \Sigma_{z_i^p})$ under H_1, where both $a_i(k-1)$ and $\Sigma_{z_i^p}$ are unknown to subsystem i. From Imhof (1961), we obtain that the detection variable $t^p(z_i^p(k))$ in (31) follows generalized Chi-squared distribution. However, the CDF of generalized Chi-squared distribution cannot be expressed in closed-form.

To cope with this problem, we consider that the subsystem $i \in S$, can estimate the unknown covariance $\Sigma_{z_i^p}$ by secondary data. Then each subsystem can construct detector based on the estimated covariance. We suppose that a set of secondary residual data $z_i^s \triangleq \{z_i^p(k^*), k^* = -K_i^*, -K_i^*+1, \ldots, -1 | z_i^p(k^*) \sim \mathcal{N}(0, \Sigma_{z_i^p})\}$ with $K_i^* > m_i$ the number of secondary data, is available to subsystem i. Then the detection problem can be presented as the following binary hypothesis test

$$H_0: \begin{cases} z_i^p(k) \sim \mathcal{N}(0, \Sigma_{z_i^p}), \\ z_i^p(k^*) \sim \mathcal{N}(0, \Sigma_{z_i^p}), k^* = -K_i^*, -K_i^*+1, \ldots, -1, \end{cases}$$

and

$$H_1: \begin{cases} z_i^p(k) \sim \mathcal{N}(C_i E_i a_i(k-1), \Sigma_{z_i^p}), \\ z_i^p(k^*) \sim \mathcal{N}(0, \Sigma_{z_i^p}), k^* = -K_i^*, -K_i^*+1, \ldots, -1. \end{cases}$$

Thus, GLRT criterion can be described as

$$\frac{f(z_i^s, z_i^p(k) | \Sigma_{z_i^p}, H_0)}{\sup_{a_i(k-1)} f(z_i^s, z_i^p(k) | \Sigma_{z_i^p}, H_1)} \underset{H_1}{\overset{H_0}{\gtrless}} \tau_i', \tag{43}$$

where $f(z_i^s, z_i^p(k) | \Sigma_{z_i^p}, H_0)$ and $f(z_i^s, z_i^p(k) | \Sigma_{z_i^p}, H_1)$ are the joint probability density functions of z_i^s and $z_i^p(k)$ under hypotheses H_0 and H_1, respectively.

Under H_0, $f(z_i^s, z_i^p(k) | \Sigma_{z_i^p}, H_0)$ is given by

$$f(z_i^s, z_i^p(k) | \Sigma_{z_i^p}, H_0) = C_i(\Sigma_{z_i^p}) \exp(\psi_i^{h_0}), \tag{44}$$

where $C_i(\Sigma_{z_i^p}) = 1/[(2\pi)^{m_i} |\Sigma_{z_i^p}|]^{\frac{K_i^*+1}{2}}$ and

$$\psi_i^{h_0} = -\frac{1}{2}[z_i^p(k)]^T \Sigma_{z_i^p}^{-1} z_i^p(k) - \frac{1}{2}\sum_{k=-K_i^*}^{-1} [z_i^p(k^*)]^T \Sigma_{z_i^p}^{-1} z_i^p(k^*).$$

It follows from Raghavan, Qiu, and McLaughlin (1995) that at time k, the unknown $\Sigma_{z_i^p}$ can be estimated as

$$\hat{\Sigma}_{z_i^p}^{h_0}(k) \triangleq \frac{1}{K_i^*+1}\Big(z_i^p(k)[z_i^p(k)]^T + \sum_{k^*=-K_i^*}^{-1} z_i^p(k^*)[z_i^p(k^*)]^T\Big).$$

Substituting the estimate $\hat{\Sigma}_{z_i^p}^{h_0}(k)$ for $\Sigma_{z_i^p}$ in (44), we have

$$f(z_i^s, z_i^p(k) | \hat{\Sigma}_{z_i^p}^{h_0}(k), H_0) = C_i(\hat{\Sigma}_{z_i^p}^{h_0}(k)) \exp(\hat{\psi}_i^{h_0}), \tag{45}$$

where $C_i(\hat{\Sigma}_{z_i^p}^{h_0}(k)) = 1/[(2\pi)^{m_i} |\hat{\Sigma}_{z_i^p}^{h_0}(k)|]^{\frac{K_i^*+1}{2}}$ and

$$\hat{\psi}_i^{h_0} = -\frac{1}{2}\{[z_i^p(k)]^T [\hat{\Sigma}_{z_i^p}^{h_0}(k)]^{-1} z_i^p(k) + \sum_{k^*=-K_i^*}^{-1} [z_i^p(k^*)]^T [\hat{\Sigma}_{z_i^p}^{h_0}(k)]^{-1} z_i^p(k^*)\}.$$

It is noted that

$$[z_i^p(k)]^T [\hat{\Sigma}_{z_i^p}^{h_0}(k)]^{-1} z_i^p(k) = \mathrm{Tr}\{[\hat{\Sigma}_{z_i^p}^{h_0}(k)]^{-1} z_i^p(k)[z_i^p(k)]^T).$$

Therefore, (45) is equivalent to

$$f(z_i^s, z_i^p(k) | \hat{\Sigma}_{z_i^p}^{h_0}(k), H_0) = \frac{1}{[(2\pi e^2)^{m_i} |\hat{\Sigma}_{z_i^p}^{h_0}(k)|]^{\frac{K_i^*+1}{2}}}. \tag{46}$$

Then under H_1, the joint density function of secondary residual data is described as

$$f(z_i^s, z_i^p(k) | \Sigma_{z_i^p}, H_1) = C_i(\Sigma_{z_i^p}) \exp(\psi_i^{h_1}), \tag{47}$$

where $\psi_i^{h_1} = -\frac{1}{2}\{\sum_{k^*=-K_i^*}^{-1} [z_i^p(k^*)]^T \Sigma_{z_i^p}^{-1} z_i^p(k^*) + [z_i^p(k) - a_i(k-1)]^T \Sigma_{z_i^p}^{-1} [z_i^p(k) - a_i(k-1)]\}$. We follow Raghavan et al. (1995) and estimate $\Sigma_{z_i^p}$ under H_1 at time k as

$$\hat{\Sigma}_{z_i^p}^{h_1}(k) \triangleq \frac{1}{K_i^*+1}\Big(\sum_{k^*=-K_i^*}^{-1} z_i^p(k^*)[z_i^p(k^*)]^T\Big).$$

Moreover, setting $a_i(k-1) = z_i^p(k)$ can maximize the function (47). Then substituting the estimation $\hat{\Sigma}_{z_i^p}^{h_1}(k)$ for $\Sigma_{z_i^p}$ in (47), we have

$$f(z_i^s, z_i^p(k)|\hat{\Sigma}_{z_i^p}^{h_1}(k), H_1) = \frac{1}{[(2\pi e^2)^{m_i}|\hat{\Sigma}_{z_i^p}^{h_1}(k)|]^{\frac{K_i^*+1}{2}}}. \tag{48}$$

Thus integrating (46) and (48), we can transform the GLRT criterion in (43) with the estimated covariances $\hat{\Sigma}_{z_i^p}^{h_0}(k)$ and $\hat{\Sigma}_{z_i^p}^{h_1}(k)$ into

$$\begin{aligned}\frac{|\hat{\Sigma}_{z_i^p}^{h_1}(k)|^{\frac{K_i^*+1}{2}}}{|\hat{\Sigma}_{z_i^p}^{h_0}(k)|^{\frac{K_i^*+1}{2}}} &= \left(\frac{|\Sigma_{z_i^p}^s|}{|\Sigma_{z_i^p}^s|(1+[z_i^p(k)]^T(\Sigma_{z_i^p}^s)^{-1}z_i^p(k))}\right)^{\frac{K_i^*+1}{2}} \\ &= \left(\frac{1}{1+[z_i^p(k)]^T(\Sigma_{z_i^p}^s)^{-1}z_i^p(k)}\right)^{\frac{K_i^*+1}{2}} \underset{H_1}{\overset{H_0}{\gtrless}} \tau_i',\end{aligned}$$

where $\Sigma_{z_i^p}^s = \sum_{k^*=-K_i^*}^{-1} z_i^p(k^*)[z_i^p(k^*)]^T$. Therefore, we get the following detector

$$t^s(z_i^p(k)) \triangleq [z_i^p(k)]^T(\Sigma_{z_i^p}^s)^{-1}z_i^p(k) \underset{H_0}{\overset{H_1}{\gtrless}} \tau_i^s - 1, \tag{49}$$

where $\tau_i^s - 1$ is detection threshold that needs to be determined.

The false alarm probability and the detection probability of detector (49) can be written as

$$\begin{aligned}P_{i,f}^s &= P(t^s(z_i^p(k)) > \tau_i^s - 1|H_0) \\ &= \binom{K_i^*}{m_i-1}\left(\frac{1}{\tau_i^s}\right)^{K_i^*-m_i+1} \\ &\quad \times {}_2F_1(K_i^*-m_i+1, 1-m_i; K_i^*-m_i+2; 1/\tau_i^s)\end{aligned} \tag{50}$$

and

$$\begin{aligned}P_{i,d}^s(k) &= P(t^s(z_i^p(k)) > \tau_i^s - 1|H_1) \\ &= \int_0^{1/\tau_i^s} f_i(r)dr + \int_{1/\tau_i^s}^1 f_i(r) \\ &\quad \times \left[1 - \frac{1}{(r\tau_i^s)^{K_i^*-m_i+1}} \sum_{l=1}^{K_i^*-m_i+1} \binom{K_i^*-m_i+1}{l}\right. \\ &\quad \left. \times (r\tau_i^s - 1)^t G_l(c_i^p(k)/\tau_i^s)dr\right],\end{aligned} \tag{51}$$

respectively, where ${}_2F_1(K_i^*-m_i+1, 1-m_i; K_i^*-m_i+2; \frac{1}{\tau_i^s})$ is the Gaussian hypergeometric series given by ${}_2F_1(K_i^*-m_i+1, 1-m_i; K_i^*-m_i+2; \frac{1}{\tau_i^s}) = 1+\sum_{l=1}^{\infty} \frac{(K_i^*-m_i+1)_l^+(1-m_i)_l^+}{(K_i^*-m_i+2)_l^+} \frac{1}{(\tau_i^s)^l l!}$, $f_i(r)$ is described as $f_i(r) = \frac{K_i^*!}{(K_i^*-m_i+1)!(m_i-2)!} r^{K_i^*-m_i+1}(1-r)^{m_i-2}$ with $0 \le r \le 1$, $G_l(c_i^p(k)/\tau_i^s)$ is given by $G_l(c_i^p(k)/\tau_i^s) = e^{-c_i^p(k)/\tau_i^s}\sum_{n=0}^{l-1}\frac{(c_i^p(k)/\tau_i^s)^n}{n!}$, and $c_i^p(k)$ is given in (32).

From (50), we observe that the false alarm probability $P_{i,f}^s$ is irrelevant to the covariance $\Sigma_{z_i^p}$. Therefore, the detector (49) ensures constant false alarm rate property with respect to the covariance. Then, it can be obtained from (51) that the detection probability $P_{i,d}^s(k)$ is related to the variable $c_i^p(k)$. To analyze the effect of $c_i^p(k)$ on $P_{i,d}^s(k)$, in the following proposition, we describe the monotonic relationship between the detection probability $P_{i,d}^s(k)$ and the variable $c_i^p(k)$.

Proposition 1. *The detection probability $P_{i,d}^s(k)$ is an increasing function of $c_i^p(k)$.*

Table 1
Subsystem parameters.

m_i^a	l_i	ε_i	k_{12}	k_{23}	k_{24}	k_{34}
0.5 kg	0.1 m	0.06 m	27	40	35	53

Proof. It suffices to prove that the derivative of $P_{i,d}^s(k)$ with respect to $c_i^p(k)$ is positive. From (51), we have

$$\begin{aligned}\frac{\partial P_{i,d}^s(k)}{\partial c_i^p(k)} &= \int_{1/\tau_i^s}^1 \frac{f_i(r)}{(r\tau_i^s)^{K_i^*-m_i+1}} \sum_{l=1}^{K_i^*-m_i+1}\binom{K_i^*-m_i+1}{l} \\ &\quad \times \sum_{n=0}^{l-1}\left[\frac{(c_i^p(k))^n}{n!(\tau_i^s)^{n+1}} - \frac{(c_i^p(k))^{n-1}}{(n-1)!(\tau_i^s)^n}\right] \\ &\quad \times (r\tau_i^s-1)^t e^{-c_i^p(k)/\tau_i^s}dr, \\ &= \int_{1/\tau_i^s}^1 \frac{f_i(r)}{(r\tau_i^s)^{K_i^*-m_i+1}} \sum_{l=1}^{K_i^*-m_i+1}\binom{K_i^*-m_i+1}{l} \\ &\quad \times (r\tau_i^s-1)^t e^{-c_i^p(k)/\tau_i^s}\frac{(c_i^p(k))^{l-1}}{(n-1)!(\tau_i^s)^l}dr,\end{aligned}$$

where $1 \le r\tau_i^s \le \tau_i^s$. Thus, we get $\frac{\partial P_{i,d}^s(k)}{\partial c_i^p(k)} > 0$. ■

Remark 6. By Proposition 1, increasing the detection probability $P_{i,d}^s(k)$ requires increasing $c_i^p(k)$. Moreover, it can be obtained from (35) that if we increase the privacy noise covariance, the variable $c_i^p(k)$ will increase. Therefore, there is also a trade-off between the detection probability and the degree of privacy preservation. In addition, the privacy noise covariance can be directly obtained by solving optimization problem (36). It is noted that different from the detection threshold obtained by (33) under known privacy noise covariance, the detection threshold $\tau_i^s - 1$ in (49) is obtained by (50) under unknown privacy noise covariance.

6. Simulation

We consider a system composed of $N = 4$ subsystems, interconnected as in Fig. 2. The system is described as the linearized model of multiple pendula coupled through a spring Barboni et al. (2019). The dynamic of subsystem i is given by

$$m_i^a l_i^2 \ddot{\delta}_i = m_i^a g^c l_i \delta_i + u_i + \sum_{j\in\mathcal{N}_i} k_{ij}\varepsilon_i^2(\delta_j - \delta_i), \tag{52}$$

where δ_i, m_i^a and l_i are the displacement angle, mass, and length of the pendulum, respectively, g^c is the gravitational constant, $k_{ij} = k_{ji}$ is the spring coefficient, and ε_i is the height at which the spring is attached to pendulum i. Some parameters used in the simulation are described in Table 1. Define the state vector $x_i = [\delta_i \;\; \dot{\delta}_i]^T$. The control law is given by $u_i = \mathcal{K}_i y_i$, where $y_i = x_i + v_i$ and $\mathcal{K}_i$ is the local controller gain. Then we discretize the dynamic of each subsystem by Euler's approximation with sampling time $T_s = 0.01$ s. Moreover, the other parameters are given by $C_i = I$, $\Sigma_{w_i} = 0.001I$ and $\Sigma_{v_i} = 0.001I$.

Starting from time $k_a = 1$ s, we assume that subsystem 3 is attacked by the attacks signals

$$\eta_3(k) = 3\left(1 - e^{-0.3(kT_s - k_a)}\right)\sin\left(\frac{2}{30}\pi kT_s\right)$$

and $\gamma_3(k) = C_3 x_3^a(k)$ with $x_3^a(k) = A_3 x_3^a(k-1) + B_3\eta_3(k-1)$.

We take subsystem 2 as an example to analyze the detection performance. The preset false alarm probability threshold is $p_2^f = 0.25$. Moreover, the neighboring subsystems of subsystem 2 transmit the noisy state estimations to subsystem 2 to protect

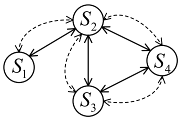
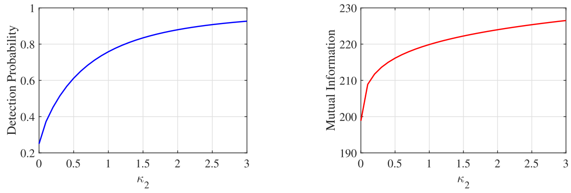


Fig. 2. Topology of interconnected system.



(a) Detection probability. (b) Mutual information.

Fig. 3. Effects of κ_2 on the mutual information and the detection probability under known privacy noise covariance.

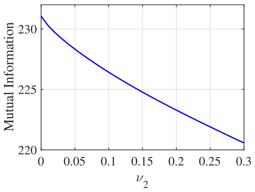


Fig. 4. Effect of ν_2 on the mutual information under unknown privacy noise covariance.

private information. Then the privacy is measured by the mutual information $I_2 \triangleq \sum_{j\in\mathcal{N}_2} I[(\tilde{x}_j)_1^K; (\theta_j^2)_1^K]$ with $K = 5$.

We first consider that the privacy noise covariance of each subsystem is known to its neighboring subsystems. By (33), we get the detection threshold $\tau_2 = 2.773$. The covariance of the privacy noise is obtained by solving optimization problem (36). Fig. 3 describes the effects of weight factor κ_2 on the mutual information I_2 and the detection probability $P_{2,d}^p(K)$. It can be seen from Fig. 3 that as the weight factor κ_2 increases, both the detection probability and the mutual information increase. Therefore, there exists a trade-off between the degree of privacy preservation and detection probability.

Then, we consider that the privacy noise covariance of each subsystem is unknown to its neighboring subsystems. To constrain the effect of privacy noise on the false alarm probability, we solve the optimization problem (42) to obtain the covariance of the privacy noise. Fig. 4 shows the effect of ν_2 on the mutual information I_2, where ν_2 is the upper bound of false alarm distortion level. It can be seen that as the false alarm probability increases, the mutual information decreases. Therefore, there exists a monotonically increasing relationship between the degree of privacy preservation and false alarm probability.

Finally, we consider the relationship between the detection probability and the mutual information under unknown privacy noise distribution. By (50), we obtain the detection threshold $\tau_3 = 3$. The privacy noise covariance is also obtained by solving the optimization problem (36). The effect of weight factor κ_2 on the mutual information I_2 is shown in Fig. 3(b). Fig. 5 shows the effect of weight factor κ_2 on the detection probability $P_{2,d}^s(K)$. It can be seen that the detection probability $P_{2,d}^s(K)$ is an increasing function of κ_2. Therefore, there also exists a trade-off between the degree of privacy preservation and the detection probability.

7. Conclusion

In this paper, we investigated the problems of attack detection and privacy preservation for interconnected system. The trade-off between privacy and security was analyzed under both known and unknown privacy noise covariance scenarios, respectively, and the corresponding optimization problems were established to obtain the privacy noise covariances. Future work may explore the trade-off between privacy and security for alternative attack detectors (e.g., CUSUM detector) and privacy preservation methods.

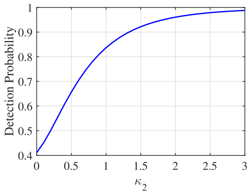


Fig. 5. Effect of κ_2 on detection probability under unknown privacy noise covariance.

References

Anguluri, R., Katewa, V., & Pasqualetti, F. (2018). Attack detection in stochastic interconnected systems: Centralized vs decentralized detectors. In *Proceedings of the 57th IEEE conference on decision and control* (pp. 4541–4546).

Barboni, A., Gallo, A. J., Boem, F., & Parisini, T. (2019). A distributed approach for the detection of covert attacks in interconnected systems with stochastic uncertainties. In *Proceedings of the 58th IEEE conference on decision and control* (pp. 5623–5628).

Barboni, A., Rezaee, H., Boem, F., & Parisini, T. (2020). Detection of covert cyber-attacks in interconnected systems: A distributed model-based approach. *IEEE Transactions on Automatic Control*, *65*(9), 3728–3741.

Boem, F., Gallo, A. J., Ferrari-Trecate, G., & Parisini, T. (2017). A distributed attack detection method for multi-agent systems governed by consensus-based control. In *Proceedings of the 56th IEEE conference on decision and control* (pp. 5961–5966).

Cover, T. M., & Thomas, J. A. (1991). Entropy, relative entropy and mutual information. *Elements of Information Theory*, *2*(1), 1–55.

Dey, K. C., Mishra, A., & Chowdhury, M. (2014). Potential of intelligent transportation systems in mitigating adverse weather impacts on road mobility: A review. *IEEE Transactions on Intelligent Transportation Systems*, *16*(3), 1107–1119.

Fang, H., & de Callafon, R. A. (2012). On the asymptotic stability of minimum-variance unbiased input and state estimation. *Automatica*, *48*(12), 3183–3186.

Farokhi, F., & Esfahani, P. M. (2018). Security versus privacy. In *Proceedings of the 57th IEEE conference on decision and control* (pp. 7101–7106).

Farokhi, F., & Sandberg, H. (2019). Ensuring privacy with constrained additive noise by minimizing Fisher information. *Automatica*, *99*, 275–288.

Ghosh, B. (1973). Some monotonicity theorems for χ^2, F and t distributions with applications. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *35*(3), 480–492.

Gillijns, S., & De Moor, B. (2007). Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica*, *43*(1), 111–116.

Han, D., Liu, K., Lin, Y., & Xia, Y. (2022). Differentially private distributed online learning over time-varying digraphs via dual averaging. *International Journal of Robust and Nonlinear Control*, *32*(5), 2485–2499.

Hayati, H., Murguia, C., & van de Wouw, N. (2024). Privacy-preserving anomaly detection in stochastic dynamical systems: Synthesis of optimal Gaussian mechanisms. *European Journal of Control*, *81*, Article 101142.

He, J., Cai, L., Cheng, P., Pan, J., & Shi, L. (2018). Distributed privacy-preserving data aggregation against dishonest nodes in network systems. *IEEE Internet of Things Journal*, *6*(2), 1462–1470.

Horn, R. A., & Johnson, C. R. (2012). *Matrix analysis*. Cambridge University Press.

Imhof, J.-P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, *48*(3/4), 419–426.

Katewa, V., Anguluri, R., & Pasqualetti, F. (2021). On a security vs privacy trade-off in interconnected dynamical systems. *Automatica*, *125*, Article 109426.

Kay, S. M. (1993). *Fundamentals of statistical signal processing: Estimation theory*. Prentice Hall: Englewood Cliffs.

Le Ny, J., & Pappas, G. J. (2013). Differentially private filtering. *IEEE Transactions on Automatic Control*, *59*(2), 341–354.

Lin, Y., Liu, K., Han, D., & Xia, Y. (2024). Statistical privacy-preserving online distributed Nash equilibrium tracking in aggregative games. *IEEE Transactions on Automatic Control*, *69*(1), 323–330.

Liu, S., Chen, B., Zourntos, T., Kundur, D., & Butler-Purry, K. (2014). A coordinated multi-switch attack for cascading failures in smart grid. *IEEE Transactions on Smart Grid*, *5*(3), 1183–1195.

Liu, Y., Xu, B., & Ding, Y. (2016). Convergence analysis of cooperative braking control for interconnected vehicle systems. *IEEE Transactions on Intelligent Transportation Systems*, *18*(7), 1894–1906.

Lu, Y., & Zhu, M. (2018). Privacy preserving distributed optimization using homomorphic encryption. *Automatica*, *96*, 314–325.

Murguia, C., Shames, I., Farokhi, F., Nešić, D., & Poor, H. V. (2021). On privacy of dynamical systems: An optimal probabilistic mapping approach. *IEEE Transactions on Information Forensics and Security*, *16*, 2608–2620.

Nekouei, E., Skoglund, M., & Johansson, K. H. (2018). Privacy of information sharing schemes in a cloud-based multi-sensor estimation problem. In *Proceedings of the American control conference* (pp. 998–1002).

Raghavan, R., Qiu, H., & McLaughlin, D. (1995). CFAR detection in clutter with unknown correlation properties. *IEEE Transactions on Aerospace and Electronic Systems*, *31*(2), 647–657.

Ruan, M., Gao, H., & Wang, Y. (2019). Secure and privacy-preserving consensus. *IEEE Transactions on Automatic Control*, *64*(10), 4035–4049.

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

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

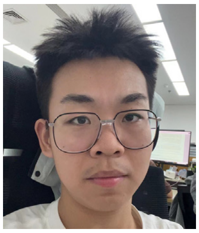
Wang, H., Liu, K., Li, B., Fridman, E., & Xia, Y. (2024). Privacy and security trade-off in interconnected systems with known or unknown privacy noise covariance. arXiv preprint arXiv:2405.16905.

Yazdani, K., Jones, A., Leahy, K., & Hale, M. (2022). Differentially private LQ control. *IEEE Transactions on Automatic Control*, *68*(2), 1061–1068.

Haojun Wang received the B.E. degree in measurement and control technology and instrument from Qingdao University of Science and Technology, in 2017, and the M.E. degree in traffic information engineering and control from Beijing Jiaotong University, in 2020. He is currently working toward the Ph.D degree in control science and engineering from Beijing Institute of Technology. His research interests include online learning, privacy preservation, and the security of cyber–physical systems.

Kun Liu received the Ph.D. degree in electrical engineering and systems from Tel Aviv University, Tel Aviv-Yafo, Israel, in 2012. From 2013 to 2015, he was a Postdoctoral Researcher with the ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden. In 2015, he held Researcher, Visiting, and Research Associate positions with, respectively, the KTH Royal Institute of Technology; CNRS, Laboratory for Analysis and Architecture of Systems, Toulouse, France; and the University of Hong Kong, Hong Kong. In 2018, he was a Visiting Scholar with INRIA, Lille, France. In 2015, he joined the School of Automation, Beijing Institute of Technology, Beijing, China. His current research interests include networked control, game-theoretic control, and security and privacy of cyber–physical systems, with applications in autonomous systems.

Dr. Liu currently serves as an Associate Editor for the IMA Journal of Mathematical Control and Information and the Journal of Beijing Institute of Technology. He is a Conference Editorial Board Member of the IEEE Control Systems Society.



Baojia Li received the B.E. degree in Automation from Central South University, in 2021, and the M.E. degree in Control Science and Engineering from Beijing Institute of Technology, in 2024. He is currently a Ph.D. candidate at the School of Automation, Beijing Institute of Technology. His research interests include security and privacy issues in cyber–physical systems and distributed optimization.

Emilia Fridman received the M.Sc. and Ph.D in mathematics in Russia. Since 1993 she has been at Tel Aviv University, where she is currently Professor in the Department of Electrical Engineering — Systems. She has held numerous visiting positions in Europe, China and Australia. Her research interests include time-delay systems, networked control systems, distributed parameter systems, robust control and extremum seeking. She has published more than 200 journal articles and 2 monographs. She serves/served as Associate Editor in Automatica, SIAM Journal on Control and Optimization and IMA Journal of Mathematical Control and Information. She is IEEE Fellow and was a member of the IFAC Council. In 2014 she was ranked as a Highly Cited Researcher by Thomson ISI. Since 2018, she has been the incumbent for Chana and Heinrich Manderman Chair on System Control at Tel Aviv University. In 2021 she was recipient of IFAC Delay Systems Life Time Achievement Award and of Kadar Award for outstanding research in Tel Aviv University. She is currently IEEE CSS Distinguished Lecturer. In 2023 her monograph "Introduction to Time-Delay Systems: Analysis and Control" (Birkhauser, 2014) was the winner of IFAC Harold Chestnut Control Engineering Textbook Prize.

Yuanqing Xia received the M.S. degree in fundamental mathematics from Anhui University, Hefei, China, in 1998, and the Ph.D. degree in control theory and control engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2001. From 2002 to 2003, he was a Post-Doctoral Research Associate with the Institute of Systems Science, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing. From 2003 to 2004, he was with the National University of Singapore, Singapore, as a Research Fellow, where he researched on variable structure control. From 2004 to 2006, he was with the University of Glamorgan, Pontypridd, U.K., as a Research Fellow. From 2007 to 2008, he was a Guest Professor with Innsbruck Medical University, Innsbruck, Austria. Since 2004, he has been with the School of Automation, Beijing Institute of Technology, Beijing, first as an Associate Professor and, then, a Professor since 2008.

His current research interests include networked control systems, robust control and signal processing, active disturbance rejection control, and flight control.