Quantification of Distributed Secrecy Loss in Stochastic Discrete Event Systems under Bounded-Delay Communications

Mariam Ibrahim[1&2], *Member, IEEE*, Jun Chen[1], *Member, IEEE* and Ratnesh Kumar[1], *Fellow, IEEE*
[1] Iowa State University, Dept. of Elec. & Comp. Eng., Ames, IA 50011.
[2] German Jordanian University, Dept. of Mechatronics Eng., Amman 11180, Jordan.
Emails: {mariami,rkumar}@iastate.edu, jchenee2015@gmail.com.

***Abstract*— Unlike information, behaviors cannot be encrypted and may instead be protected by providing covers that generate indistinguishable observations from behaviors needed to be kept secret. Such a scheme may still leak information about secrets due to statistical difference between the occurrence probabilities of the secrets and their covers. Jensen-Shannon Divergence (JSD) is a possible means of quantifying statistical difference between two distributions and was used to measure such information leak as in our earlier work [1]. This paper studies secrecy quantification in stochastic partially-observed discrete event systems in the presence of distributed collusive attackers/observers, each with its own local partial observability, generalizing the setting of single observer in [1]. The local observers collude and exchange their observations over communication channels that introduce bounded delays. We propose a method to compute the JSD-based secrecy measure in this distributed setting by introducing bounded-delay channel models to extend the system model to capture the effect of exchange of observations, and to measure the distributed secrecy loss.**

I. Introduction

Growing progress in information and communication technologies has led to growth in eavesdropping and tampering of private communication or behaviors. In contrast to information, behaviors cannot be encrypted, and their *secrecy* can instead be attained through introduction of *covers* that ambiguate secrets in presence of partial observation. Many techniques for hiding secrets based on ambiguation schemes have been proposed as, *Steganography and Watermarking*, *Network level Anonymization*, and *Software Obfuscation*.

Also, various notions of information secrecy have also been explored in literature. For example, [2] examines non-interference, requiring that secrets (private variables) do not interfere with or influence the observables (public variables). Non-interference is a logical notion, but for stochastic systems, the mutual information between the private and public variables can be used to quantify the level of interference, and hence loss of secrecy [2]. Mutual information is only an average case measure, and a worst case measure can also be defined, using for example *min-entropy* [3]. Extension of the notion of non-interference over behaviors (sequences) was explored in [4], requiring that every secret behavior must be masked by a cover behavior.

For information leakage over *sequences of observations* from a stochastic systems, mutual information can again be used to quantify the level of secrecy loss, and as shown in [1], it can be related to a certain Jensen-Shannon Divergence (JSD) computation, and can be used to measure the disparity between the distributions of a secret versus its cover as a way to quantify the secrecy [5]. In a similar spirit, [6] considered mutual information between the secret states and the observed behaviors, and required it to be upper bounded. Checking this is undecidable, and [7] proposed a stronger notion, requiring the probability of revealing secrets to remain upper bounded at each time step. In contrast, S_τ-secrecy [8] bounds the probability of revealing secrets over the set of *all* behaviors, as opposed to for each step. More related works on secrecy can be found in a recent survey [9].

The work presented here extends our prior work on secrecy quantification for partially-observed discrete event systems (PODESs) in [1] from a centralized single observer/attacker setting to a distributed and collusive setting of multiple observers/attackers that have their own personal observations, and also collude by exchanging their observations over channels, that introduce delays that are bounded. Our goal then is to quantify the level of loss of secrecy in such a distributed collusive setting. We accomplish this by introducing channel models and extending the system model by including in it all incoming channel models as in [10]. We then compute the JSD-based secrecy measure with respect to the extended model.

II. Notation and Preliminaries

A. Stochastic PODESs

For an event set Σ, define $\overline{\Sigma} := \Sigma \cup \{\epsilon\}$, where ϵ denotes "no-event". The set of all finite length event sequences over Σ, including ϵ is denoted as Σ^*, $\Sigma^+ := \Sigma^* - \{\epsilon\}$, and Σ^n is the set of event sequences of length $n \in \mathbb{N}$. A *trace* is a member of Σ^* and a *language* is a subset of Σ^*. We use $s \leq t$ to denote if $s \in \Sigma^*$ is a prefix of $t \in \Sigma^*$, and $|s|$ to denote the length of s or the number of events in s. For $L \subseteq \Sigma^*$, its prefix-closure is defined as $pr(L) := \{s \in \Sigma^* | \exists t \in \Sigma^* : st \in L\}$ and L is said to be prefix-closed (or simply closed) if $pr(L) = L$, i.e., whenever L contains a trace, it also contains all the prefixes of that trace. For $s \in \Sigma^*$ and $L \subseteq \Sigma^*$, $L\backslash s := \{t \in \Sigma^* | st \in L\}$ denotes the set of traces in L *after* s.

A stochastic PODES can be modeled by a *stochastic automaton* $G = (X, \Sigma, \alpha, x_0)$, where X is the set of states, Σ is the finite set of events, $x_0 \in X$ is the initial state, and $\alpha : X \times \Sigma \times X \to [0, 1]$ is the probability transition function [11], and $\forall x \in X, \sum_{\sigma \in \Sigma} \sum_{x' \in X} \alpha(x, \sigma, x') = 1$. A non-stochastic PODES can be modeled as the same 4-tuple, but by replacing the transition function with $\alpha : X \times \Sigma \times X \to \{0, 1\}$, and a non-stochastic DES is deterministic if $\forall x \in X, \sigma \in \Sigma, \sum_{x' \in X} \alpha(x, \sigma, x') \in \{0, 1\}$. The transition probability function α can be generalized to $\alpha : X \times \Sigma^* \times X$ in a natural way: $\forall x_i, x_j \in X, s \in \Sigma^*, \sigma \in \Sigma, \alpha(x_i, s\sigma, x_j) = \sum_{x_k \in X} \alpha(x_i, s, x_k)\alpha(x_k, \sigma, x_j)$, and $\alpha(x_i, \epsilon, x_j) = 1$ if $x_i = x_j$ and 0 otherwise. Define the language generated by G as $L(G) := \{s \in \Sigma^* \mid \exists x \in X, \alpha(x_0, s, x) > 0\}$. For a given G, a *component* $C = (X_C, \alpha_C)$ of G is a "subgraph" of G, i.e., $X_C \subseteq X$ and $\forall x, x' \in X_C$ and $\sigma \in \Sigma$, $\alpha_C(x, \sigma, x') = \alpha(x, \sigma, x')$ whenever the latter

This research was supported in part by Security and Software Engineering Research Center (S2ERC), and the National Science Foundation under the grants NSF-CCF-1331390 and NSF-ECCS 1509420. [1] Jun Chen is now with Idaho National Laboratory upon graduation from ISU.

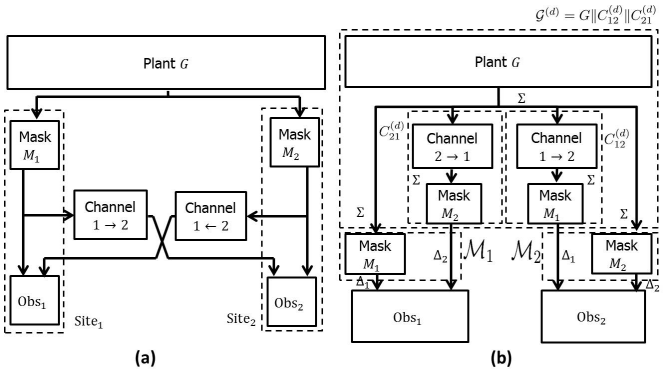

Fig. 1. (a) Distributed secrecy system architecture to (b) equivalent system architecture.

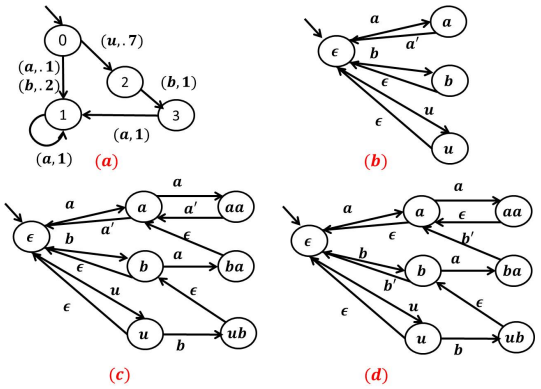

Fig. 2. (a) Stochastic PODES G; (b) $C_{12}^{(0)}$; (c) $C_{12}^{(1)}$; (d) $C_{21}^{(1)}$.

is positive, and $\alpha_C(x, \sigma, x') = 0$ otherwise. C is said to be a *strongly connected component* (SCC) or *irreducible* if $\forall x, x' \in X_C$, $\exists s \in \Sigma^*$ such that $\alpha_C(x, s, x') > 0$. A SCC C is said to be *closed* if for each $x \in X_C$, $\sum_{\sigma \in \Sigma} \sum_{x' \in X_C} \alpha_C(x, \sigma, x') = 1$. The states which belong to a closed SCC are *recurrent states* and the remaining states (that do not belong to any closed SCC) are *transient states*. Another way to identify recurrent versus transient states is to consider the steady-state state distribution π^* as the fixed-point of $\pi^* = \pi^* \Omega$, where π^* is a row-vector with the same size as X, and Ω is the transition matrix with ijth entry being the transition probability $\sum_{\sigma \in \Sigma} \alpha(i, \sigma, j)$. (In case Ω is periodic with period $d \neq 1$, we consider the set of fixed-points of $\pi^* = \pi^* \Omega^d$). Then any state i is recurrent if and only if there exists a reachable fixed point π^* such that the ith entry of π^* is nonzero. Identifying the set of recurrent states can be done polynomially, by the algorithm presented in [12].

B. d-delaying&masking communication channel

Fig. 1(a) shows the architecture of a system with distributed observers/attackers, where it is assumed for simplicity and without loss of any generality that there are two local observers at two local sites $I = \{1, 2\}$. Each site has three modules [10]: (i) observation mask $M_i : \overline{\Sigma} \to \overline{\Delta}_i$, where Δ_i is the set of locally observed symbols and $M_i(\epsilon) = \epsilon$ (M_i can be extended to Σ^* as follows: $M_i(\epsilon) = \epsilon$, and $\forall s \in \Sigma^*, \sigma \in \overline{\Sigma}, M_i(s\sigma) = M_i(s) M_i(\sigma)$), (ii) communication channels $C_{ij}^{(d)}, j \neq i$, $i, j \in I$, which are lossless and order-preserving, but introduce delays bounded by d, and (iii) observer Obs_i, that tracks the system "information-state" following the arrival of its local observations and the communicated observations received from other sites $j \in I, j \neq i$.

The communication channel is a "*delay-block*" with d-bounded communication delay that holds the transmitted information in First-In-First-Out ($FIFO$) manner for at most d delay steps. Accordingly, since there can be at most d events executed by system G between the transmission and the reception of a message on a channel, the channel has a maximum queue length $d + 1$. Also, the channel queue evolves whenever a system event occurs, or a transmitted observation is delivered to a destination observer, where such arrival and departure events occur asynchronously. Accordingly, the d-delaying&masking non-stochastic channel model from site-i to site-j ($i \neq j$, $i, j \in I$) is of the form, $C_{ij}^{(d)} = (Q_{ij}^{(d)}, \Sigma \cup \overline{\Delta}_i, \beta_{ij}^{(d)}, q_0)$, with the elements as follows. $Q_{ij}^{(d)} \subseteq \Sigma^*$ denotes the set of states, which are the event traces executed in the system but their observed values pending to be delivered at the destination. For $q \in Q_{ij}^{(d)}$, it holds that $|q| \leq d+1$. $\Sigma \cup \overline{\Delta}_i$ is the event set of $C_{ij}^{(d)}$, where Σ is its set of input events and $\overline{\Delta}_i$ is its set of output events. Without loss of generality, we assume that $\Sigma \cap \overline{\Delta}_i = \emptyset$, and $\Delta_i \cap \Delta_j = \emptyset, (j \neq i)$ (otherwise, we can simply rename some of the symbols). $q_0 = \epsilon$ is the initial state, whereas the transition function $\beta_{ij}^{(d)}$ is defined as follows:

1) "Arrival" due to an event execution in the system: $\forall q \in Q_{ij}^{(d)}, \forall \sigma \in \Sigma$, if $|q| \leq d$, then $\beta_{ij}^{(d)}(q, \sigma) = q\sigma$,
2) "Departure" due to a reception at the destination observer: $\forall q \in Q_{ij}^{(d)}, \forall \delta_i \in \overline{\Delta}_i$, if $M_i(head(q)) = \delta_i$, then $\beta_{ij}^{(d)}(q, \delta_i) = q \backslash head(q)$,
3) Undefined, otherwise,

where $head(q)$ is the first event in trace q, and the after operator "$\backslash$" in $q \backslash head(q)$ returns the trace after removing the initial event $head(q)$ from the trace q.

Example 1: A system model G is shown in Fig. 2(a), with $L(G) = a^+ \cup ba^* \cup uba^+$. Suppose the observation masks of two local sites are defined as follows:

- $M_1(a) = a'$, $M_1(b) = M_1(u) = \epsilon$, and
- $M_2(b) = b'$, $M_2(a) = M_2(u) = \epsilon$.

For delay $d = 0$, Fig. 2(b) shows the model $C_{12}^{(0)}$, and for delay $d = 1$, Fig. 2(c) and Fig. 2(d) show the models $C_{12}^{(1)}$ and $C_{21}^{(1)}$, respectively. If we follow the trace bab' in $C_{21}^{(1)}$, the states ϵ, b, ba and a are traversed sequentially. This corresponds to the situation in which site-2 sends out its observation b' to site-1 following the execution of ba in the system, whereas the observation of event a is pending to be received at site-1.

Next, since the operations of masking and delaying can be interchanged, the behaviors under the schematic of Fig. 1(a) are equivalent to those of Fig. 1(b). Then, it is clear that the distributed setting of Fig. 1(a) can be converted to a decentralized setting of Fig. 1(b), having an extended system $\mathcal{G}^{(d)}$ and local observers having the extended observation masks $\{\mathcal{M}_i\}$, defined below. The extended system is given by $\mathcal{G}^{(d)} = G \|_{i,j \in I, i \neq j} C_{ij}^{(d)}$, whereas the extended system model $\mathcal{G}_i$ at site-i ($i \in I$) includes the system model and only the incoming channel models: $\mathcal{G}_i = G \|_{j \in I - \{i\}} C_{ji}^{(d)}$. The *extended system* $\mathcal{G}_i$ "generates" events in $\overline{\Sigma} \cup_{j \neq i} \Delta_j$, which are observed by site-i observer Obs_i through an extended observation mask $\mathcal{M}_i : \overline{\Sigma} \cup_{j \in I - \{i\}} \overline{\Delta}_j \to \overline{\Delta} = \cup_{i \in I} \overline{\Delta}_i$. $\mathcal{M}_i$ acts the same as M_i for events in Σ, whereas it is an identity mask for events in Δ_j ($j \neq i$). Formally, it is defined

as follows: $\mathcal{M}_i(\sigma) := \begin{cases} M_i(\sigma), & \sigma \in \Sigma, \\ \sigma, & \sigma \in \Delta_j \ (j \neq i). \end{cases}$

C. Secret/non-secret behaviors and refined extended plant

Certain plant behaviors may be considered sensitive and hence secret, whereas the remaining behaviors act as cover for the secrets. Letting $L = L(G)$, suppose $K \subset L$ models the secret behaviors (also called specifications), while the remaining traces in $L - K$ act as its cover. K may be modeled by a deterministic acceptor $R = (Y, \Sigma, \beta, y_0)$ such that $L(R) = K$. By introducing a dump state D in R, and completing its transition function, we can obtain $\overline{R} = (\overline{Y}, \Sigma, \overline{\beta}, y_0)$, where $\overline{Y} = Y \cup D$, and $\forall \overline{y}, \overline{y}' \in \overline{Y}, \sigma \in \Sigma$,

$$\overline{\beta}(\overline{y}, \sigma, \overline{y}') := \begin{cases} \beta(\overline{y}, \sigma, \overline{y}') \text{ if } (\overline{y}, \overline{y}' \in Y) \wedge (\beta(\overline{y}, \sigma, \overline{y}') > 0), \\ 1 \text{ if } [(\overline{y} = \overline{y}' = D) \\ \quad \vee (\overline{y}' = D \wedge \sum_{y \in Y} \beta(\overline{y}, \sigma, y) = 0)]. \end{cases}$$

Then, the extended plant model at site-i ($i \in I$) can be refined with respect to the specification to identify the secret and cover behaviors as *states* in the refined plant, and is given by $\mathcal{G}_i^R = G||_{j \in I - \{i\}} C_{ji}^{(d)} || \overline{R}$.

Next, we assign probabilities to transitions in $\mathcal{G}_i^R$ as follows. For each state in $\mathcal{G}_i^R$, the transition is either one of the system events, or at most one of channel j ($j \neq i$) events (either arrival or departure of that channel). Suppose at a system $\mathcal{G}_i^R$ state, with vector of all incoming channel lengths $\vec{k}$, the system event is picked with probability $p_{\vec{k}}^0$, and suppose the channel j ($j \neq i$) event can occur with probability $p_{\vec{k}}^j$ such that, $p_{\vec{k}}^0 + \sum_{j \neq i} p_{\vec{k}}^j = 1$. We also require that when all channels are empty ($\vec{k} = \vec{0}$), $p_{\vec{k}}^0 = 1$ (so no channel output can occur when channels are empty), when all channels are full ($\vec{k} = \overrightarrow{d+1}$), $p_{\vec{k}}^0 = 0$ (so no channel input can occur when channels are full), and if channel j has higher queue length than channel j' ($\vec{k_j} \geq \vec{k_j'}$), then it can be expected that $p_{\vec{k}}^j \geq p_{\vec{k}}^{j'}$ (channel j event is more likely than channel j' event when channel j has more number of pending observations). With this choice of selection probability of events, refined extended system model at each site-i is given by a product of the plant model and all the incoming channel models: $\mathcal{G}_i^R = (X \times (\Pi_{j \neq i} Q_{ji}^{(d)}) \times \overline{Y}, \Sigma \cup_{j \neq i} \overline{\Delta}_j, \gamma, (x_0, \vec{q_0}, y_0))$, where $\overline{Y} = Y \cup \{D\}$, and $\forall (x, \vec{q}, \overline{y}), (x', \vec{q'}, \overline{y}') \in X \times (\Pi_{j \neq i} Q_{ji}^{(d)}) \times \overline{Y}, \sigma \in \Sigma \cup_{j \neq i} \overline{\Delta}_j$,

$$\gamma((x, \vec{q}, \overline{y}), \sigma, (x', \vec{q'}, \overline{y}')) = \begin{cases} \alpha(x, \sigma, x') \times p_{\vec{k}}^0 & \text{if } \sigma \in \Sigma, \\ p_{\vec{k}}^j & \text{if } \sigma \in \cup_{j \neq i} \overline{\Delta}_j, \end{cases}$$

if the following holds: $(\overline{y}, \overline{y}' \in Y \wedge \beta(\overline{y}, \sigma, \overline{y}') > 0) \vee (\overline{y} = \overline{y}' = D) \vee (\overline{y}' = D \wedge \sum_{y \in Y} \beta(\overline{y}, \sigma, y) = 0)$, and otherwise, $\gamma((x, \vec{q}, \overline{y}), \sigma, (x', \vec{q'}, \overline{y}')) = 0$.

The appendix describes the computation of an observer transition structure for $\mathcal{G}_i^R$ that can be used to track its evolution over only its observed symbols $\Delta := \cup_{i \in I} \Delta_i$, and also the associated transition matrices $\{\Theta(\delta) \mid \delta \in \Delta\}$.

Example 2: Continuing Example 1, suppose the delay bound $d = 1$, so there are three possibilities for the length of the only channel, $\vec{k} = \{0, 1, 2\}$. Let $p_0^0 = 1$, $p_1^0 = 0.5$, $p_2^0 = 0$ (implying $p_0^2 = 1 - p_0^0 = 0$, $p_1^2 = 1 - p_1^0 = 0.5$, $p_2^2 = 1 - p_2^0 = 1$). Fig. 3 shows the extended plant model $\mathcal{G}_1$ at site-1. Suppose R is given in Fig. 4(a), i.e., $K = L(R) = a^+ \cup ba^*$. Then, the refinement $\mathcal{G}_1^R$ is shown in Fig. 4(b). So for example, at the initial state $(0, \epsilon, 0)$, the channel is empty, and no channel events occur at this state

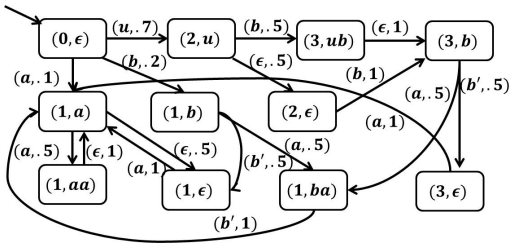

Fig. 3. Extended plant model $\mathcal{G}_1$ at site-1.

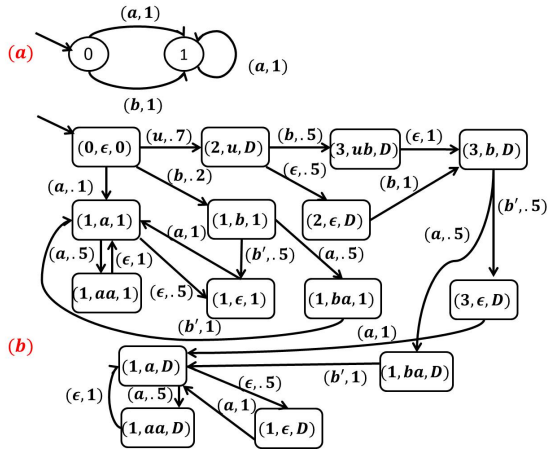

Fig. 4. (a) Specification for secrets R; (b) refined plant model $\mathcal{G}_1^R$.

($p_0^2 = 0$ while $p_0^0 = 1$). Then, for any plant event $\sigma \in \Sigma$, $\gamma((0, \epsilon, 0), u, (2, u, D)) = \alpha(0, u, 2) \times p_0^0 = 0.7 \times 1 = 0.7$, $\gamma((0, \epsilon, 0), b, (1, b, 1)) = \alpha(0, b, 1) \times p_0^0 = 0.2 \times 1 = 0.2$, and $\gamma((0, \epsilon, 0), a, (1, a, 1)) = \alpha(0, a, 1) \times p_0^0 = 0.1 \times 1 = 0.1$. Whereas, at state $(2, u, D)$, there is observation u queued up in the channel. Thus, either the plant can execute a new event $b \in \Sigma$, with probability $\gamma((2, u, D), b, (3, ub, D)) = \alpha(2, b, 3) \times p_1^0 = 1 \times p_1^0 = 0.5$, or a channel event can occur, with probability $\gamma((2, u, D), \epsilon, (2, \epsilon, D)) = p_1^2 = 0.5$. The remaining state transitions can be computed similarly. The models $\mathcal{G}_2$ and $\mathcal{G}_2^R$ at site-2 can be generated in a manner similar to $\mathcal{G}_1$ and $\mathcal{G}_1^R$, respectively.

III. Jensen-Shannon Divergence Based Distributed Secrecy Quantification

In in our earlier work [1], we presented a way to compute JSD-based measure of secrecy for stochastic PODES when there is a single observer by computing the JSD between the conditional distributions of secrets versus covers over observations of common length, and its "limiting" value as this common length approaches infinity. The JSD-based measure can continue to be used for secrecy loss quantification in the distributed collusive setting. To compute the secrecy loss in the distributed setting, resulting from the aggregated observations at any site-i ($i \in I$), which include it's own immediate observations and the delayed communicated observations from other distributed sites, the JSD computation can be carried out over the refined extended system model $\mathcal{G}_i^R$, following the method introduced in [1], and repeated below for the sake of completeness.

We begin by summarizing some relevant information theoretic notations. Given a probability distribution p over discrete set A, the entropy of p is defined as $H(p) = -\sum_{a \in A} p(a) \log p(a)$. Given two probability distributions p and q over A, the Kullback-Leibler (KL) divergences between p and q, denoted as $D_{KL}(p, q)$, is defined as

$D_{KL}(p,q) = \sum_{a\in A} p(a)\log\frac{p(a)}{q(a)}$. Given $\lambda_1 > 0$ and $\lambda_2 > 0$ satisfying $\lambda_1 + \lambda_2 = 1$, the Jensen-Shannon Divergence (JSD) between p and q under the weights (λ_1, λ_2), is defined as $D(p,q) = \lambda_1 D_{KL}(p, \lambda_1 p + \lambda_2 q) + \lambda_2 D_{KL}(q, \lambda_1 p + \lambda_2 q)$, which is equivalent to $D(p,q) = H(\lambda_1 p + \lambda_2 q) - \lambda_1 H(p) - \lambda_2 H(q)$ (for more details, refer to [13], [14]). Given two probability distributions p over A and q over B, the mutual information between p and q is defined as $I(p,q) = \sum_{a\in A, b\in B} Pr(a,b)\log\frac{Pr(a,b)}{p(a)q(b)}$. Mutual information can also be equivalently defined as $I(p,q) = H(p) - H(p|q)$, where the conditional entropy $H(p|q)$ is given as $H(p|q) = -\sum_{a\in A} p(a)\sum_{b\in B} Pr(b|a)\log Pr(b|a)$.

At each site $i \in I$, given a length-n aggregated observation $o \in \Delta^n$, let $p_n(o)$ denote its probability. Then, since the occurrences of observations of length n are mutually disjoint, $\sum_{o\in\Delta^n} p_n(o) = 1$, i.e., p_n is a probability distribution over Δ^n. Then we write its entropy as: $H(p_n) = -\sum_{o\in\Delta^n} p_n(o)\log p_n(o) = H(p_{n-1}) - \sum_{o\in\Delta^{n-1}} p_{n-1}(o)\sum_{\delta\in\Delta} p(\delta|o)\log p(\delta|o)$. Observations in Δ^n can be generated by secrets (behaviors in K) or by covers (behaviors in $L-K$), and so we define two more probability distributions over Δ^n: probability that an observation $o \in \Delta^n$ is generated by some secret in K, denoted $p_n^s(o)$, versus that is generated by some cover in $L-K$, denoted $p_n^c(o)$:

$$p_n^s(o) := \frac{Pr(s \in K \cap \Pi_\Sigma(\mathcal{M}_i^{-1}(o)))}{Pr(s \in K \cap \Pi_\Sigma(\mathcal{M}_i^{-1}(\Delta^n)))},$$

$$p_n^c(o) := \frac{Pr(s \in (L-K) \cap \Pi_\Sigma(\mathcal{M}_i^{-1}(o)))}{Pr(s \in (L-K) \cap \Pi_\Sigma(\mathcal{M}_i^{-1}(\Delta^n)))},$$

where $\Pi_\Sigma(\cdot)$ denotes the natural projection to the event set Σ. Further, define $\lambda_n^s := Pr(s \in K \cap \Pi_\Sigma(\mathcal{M}_i^{-1}(\Delta^n)))$ to be the probability of secrets and $\lambda_n^c := Pr(s \in (L-K) \cap \Pi_\Sigma(\mathcal{M}_i^{-1}(\Delta^n)))$ to be the probability of covers, respectively, generating length-n observation. Then it is easy to show that $\lambda_n^s + \lambda_n^c = 1$ for all $n \in \mathbb{N}$.

The ability of an intruder at site $i \in I$ to identify secret versus cover behaviors based on observations of length n, depends on the disparity between the two distributions p_n^s versus p_n^c: If p_n^s and p_n^c are identical, i.e., with "zero disparity", there is no way to statistically tell apart secrets from covers, and in that case there is perfect secrecy. However, when p_n^s and p_n^c are different, then one could characterize the ability of an intruder to discriminate secrets from covers, based on length-n observations, using the JSD between p_n^s and p_n^c under the weights $(\lambda_n^s, \lambda_n^c)$, denoted $D_i(p_n^s, p_n^c) = H(\lambda_n^s p_n^s + \lambda_n^c p_n^c) - \lambda_n^s H(p_n^s) - \lambda_n^c H(p_n^c)$.

The following theorem from [1] shows that the JSD measure is indeed a useful measure of information revealed, as it equals the mutual information between the observations p_n and the status (whether secret or cover) of system executions. This status can be captured by a bi-valued random variable Λ_n, defined for each $n \in \mathbb{N}$, such that $Pr(\Lambda_n = s) = \lambda_n^s$ and $Pr(\Lambda_n = c) = \lambda_n^c$.

Theorem 1 ([1]): The JSD between p_n^s and p_n^c equals the mutual information between Λ_n and p_n, i.e.,

$$D_i(p_n^s, p_n^c) = I(\Lambda_n, p_n).$$

At site $i \in I$, an intruder is likely to discriminate more if he/she observes for a longer period, and accordingly, our goal is to evaluate the worst-case loss of secrecy as obtain in the limit: $\lim_{n\to\infty} D_i(p_n^s, p_n^c)$. This worst-case JSD provides an upper bound to the amount of information leaked about secrets.

In order to numerically compute JSD at each site $i \in I$, [1] maps the JSD computation to a computation based on the state-distribution of the observer, following each observation. Each observation $o \in \Delta^*$ results in a conditional state distribution $\pi(o)$, which can be computed recursively as follows: for any $o \in \Delta^*, \delta \in \Delta$: $\pi(\epsilon) = \pi_0$ and $\pi(o\delta) = \frac{\pi(o)\times\Theta(\delta)}{||\pi(o)\times\Theta(\delta)||}$ [15], where π_0 is the initial state distribution, whereas the computation of transition matrix $\Theta(\delta)$ is given in the appendix. Let Π denote the set of all such conditional state distributions, and for each $\pi \in \Pi$ and $n \in \mathbb{N}$, denote $P_n(\pi) = Pr(o \in \Delta^n : \pi(o) = \pi)$, which is the probability that the set of all observations of length n, upon which the conditional state distribution is π. For a state distribution π, define the following notations:

$$\lambda^{s|\pi} := \sum_{\delta\in\Delta} \pi\Theta(\delta)\mathcal{I}^s,\ \lambda^{c|\pi} := \sum_{\delta\in\Delta} \pi\Theta(\delta)\mathcal{I}^c$$

$$p^{s|\pi}(\delta) := \frac{\pi\Theta(\delta)\mathcal{I}^s}{\lambda^{s|\pi}},\ p^{c|\pi}(\delta) := \frac{\pi\Theta(\delta)\mathcal{I}^c}{\lambda^{c|\pi}},$$

where $\mathcal{I}^s$ and $\mathcal{I}^c$ denote indicator column vectors of same size as number of states, with binary entries to identify the secret versus cover states (states reached by traces in K versus $L-K$). Then, as shown in Lemma 4 of [1],

$$D_i(p_n^s, p_n^c) = H(\{\lambda_n^s, \lambda_n^c\}) + \sum_{\pi\in\Pi} P_{n-1}(\pi)\Big[-H(\{\lambda^{s|\pi}, \lambda^{c|\pi}\}) + D_i(p^{s|\pi}, p^{c|\pi})\Big]. \quad (1)$$

In the limit when $n \to \infty$, if the distribution $P_n(\cdot)$ over Π converges to $P^*(\cdot)$, then $\lim_{n\to\infty} D_i(p_n^s, p_n^c)$ exists. See for example [16] for a condition under which such a convergence is guaranteed: It requires the system to be ergodic (period equals 1 and irreducible) and the existence of a finite sequence $e_1, \ldots, e_m$ such that $\Theta(e_1)\ldots\Theta(e_2)$ is a nonzero subrectangular matrix.

At each site $i \in I$, the computation of $\lim_{n\to\infty} D_i(p_n^s, p_n^c)$ using (1), requires the computation of $\lim_{n\to\infty} P_{n-1}(\pi)$ which can be accomplished with the help of an extended observer introduced in [1]. An extended observer tracks the possible system states following each observation, and also allows the computation of the corresponding state distribution. We let Obs_i be an extended observer automaton with state set $Z \subseteq 2^{X\times(\Pi_{j\neq i}Q_{ji}^{(d)})\times\overline{Y}}$, so that each node $z \in Z$ of the observer is a subset of the refined extended system states, i.e., $z \subseteq (X, (\Pi_{j\neq i}Q_{ji}^{(d)}), \overline{Y})$, and we use $|z|$ to denote the number of system states in z. Obs_i is initialized at node $z_0 = \{(x_0, \vec{q_0}, y_0)\}$, and there is a transition labeled with $\delta \in \Delta$ from node z to z' if and only if every element of z' is reachable from some elements of z along a trace that ends in the only observation δ, i.e., $z' = \{(x', \vec{q'}, \overline{y}') \in X \times (\Pi_{j\neq i}Q_{ji}^{(d)}) \times \overline{Y} : \exists(x, \vec{q}, \overline{y}) \in z, L_{\mathcal{G}_i^R}((x, \vec{q}, \overline{y}), \delta, (x', \vec{q'}, \overline{y}')) \neq \emptyset\}$. Associated with this transition is the transition probability matrix $\Theta_{z,\delta,z'}$ of size $|z|$ by $|z'|$ (a submatrix of $\Theta(\delta)$ matrix given in the appendix), whose ijth element is $\theta_{i,\delta,j}$, which is the transition probability from ith element $(x, \vec{q}, \overline{y})$ of z to jth element $(x', \vec{q'}, \overline{y}')$ of z' while producing the observation δ, and equals $\alpha(L_{\mathcal{G}_i^R}((x, \vec{q}, \overline{y}), \delta, (x', \vec{q'}, \overline{y}')))$.

Example 3: Consider the refined extended plant model of Fig. 4(b) at site-1, where $\mathcal{M}_1(a) = a'$, $\mathcal{M}_1(b) = \mathcal{M}_1(u) = \epsilon$, while the extended mask function is the identity function over the received observations, $\Delta_2 = \{b'\}$. Then, Fig. 5 shows the extended observer Obs_1.

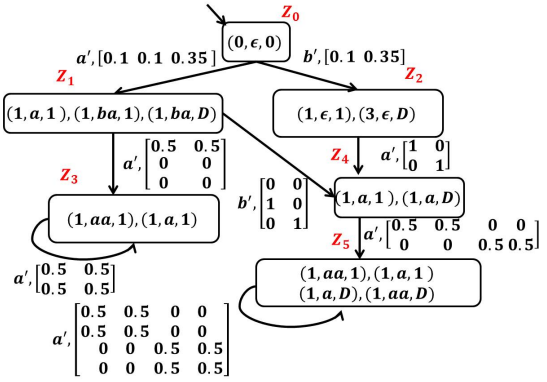

Fig. 5. Observer Obs_1 for the system of Fig. 4(b).

Associated with each observation $o \in \Delta^*$, there is a reachable state distribution $\pi(o)$ as discussed earlier. Let the state z be reached in Obs_i following observation o. Then, obviously the number of positive elements of $\pi(o)$ is the same as the number of elements in z. Then with a slight abuse of notation, we also use $\pi(o)$ to denote the row-vector containing only positive elements, and of same size as the number of elements in the node reached by o in Obs_i. Then $\pi(o)$ can also be recursively computed as follows: for any $o \in \Delta^*, \delta \in \Delta$: $\pi(\epsilon) = 1$ and $\pi(o\delta) = \frac{\pi(o)\times\Theta_{z_o,\delta,z_{o\delta}}}{||\pi(o)\times\Theta_{z_o,\delta,z_{o\delta}}||}$, where z_o and $z_{o\delta}$ are the nodes reached in Obs_i following o and $o\delta$ respectively. Then it can be seen that along any cycle in Obs_i, the distribution upon completing the cycle is a function of the distribution upon entering the cycle, through a sequence of transition matrix-multiplications and their normalizations. In case of steady-state, those two distributions will be the same, namely, a fixed point of that function. The following assumption is made as in [1].

Assumption 1 ([1]): Assume that for any sufficiently long observations $o_1 \leq o_2$, if Obs_i reaches the same node following o_1 and o_2, then $\pi(o_1) = \pi(o_2)$.

The as shown in [1], the following procedure computes the worst-case loss of secrecy at site $i \in I$, $\lim_{n\to\infty} D_i(p_n^s, p_n^c)$, under Assumption 1.

Algorithm 1:

1) Construct a $(\sum_z |z|)\times(\sum_z |z|)$ square matrix $\widetilde{\Theta}$, whose ijth block is the $|z_i|\times|z_j|$ matrix $\sum_\delta \Theta_{z_i,\delta,z_j}$. Compute the fix point distribution associated with $\widetilde{\Theta}$ by solving $\pi^* = \pi^*\widetilde{\Theta}$, where π^* is a row vector of size $\sum_z |z|$. For each $z_i \in Z$, let $p(z_i)$ be the summation of the ith block of π^*, then z_i is *recurrent* if $p(z_i) > 0$. Also note that for each $z \in Z$, exists a sufficiently large N such that $p(z) = \sum_{o\in\Delta^N : o \text{ reaches } z} p_N(o)$. In other words, $p(z)$ computes the probability of all sufficiently long observations that reach the observer state z.
2) Obtain λ^s as the summation of the elements of π^* corresponding to the secret states, i.e., $\lambda^s := \pi^*\mathcal{I}^s$, and $\lambda^c = 1 - \lambda^s$.
3) For a set of recurrent nodes $\{z_1, z_2, \ldots, z_n\}$ that form a SCC, define a set of distributions $\{\pi^*_{z_1}, \pi^*_{z_2}, \ldots, \pi^*_{z_n}\}$ to be a set of steady state distributions if $\forall i, j, \delta$, such that Θ_{z_i,δ,z_j} is defined, the following holds: $\pi^*_{z_j} = \frac{\pi^*_{z_i}\Theta_{z_i,\delta,z_j}}{||\pi^*_{z_i}\Theta_{z_i,\delta,z_j}||}$, i.e., $\pi^*_{z_i}$ represents a steady state conditional distribution following a single sufficiently long observation, that reaches z_i. Note that in this case, any other extension of o that also reaches z_i will induce the same conditional distribution $\pi^*_{z_i}$. There may exist multiple sets of steady state distributions for a given set of recurrent nodes, denoted say as $\{\{\pi^*_{z_1,k}, \ldots, \pi^*_{z_n,k}\}, k \in \mathbb{N}\}$. Then, if steady-state always exists, for any sufficiently long observation that reaches a recurrent node z, there exists $k \in \mathbb{N}$ such that $\pi(o) = \pi^*_{z,k}$. Denote $p(z,k) := Pr[\{o \mid o \text{ reaches } z \text{ and } \pi(o) = \pi^*_{z,k}\}]$.
4) Let $\mathcal{I}^s_{z'}$ and $\mathcal{I}^c_{z'}$ be indicator column vectors with binary entries of size $|z'|$ for identifying within z', the secret and cover states, respectively. For each steady state distribution $\pi^*_{z,k}$ of each recurrent node z, define:
$$\lambda^{s|\pi^*_{z,k}} := \sum_{\delta\in\Delta} \pi^*_{z,k}\Theta_{z,\delta,z'}\mathcal{I}^s_{z'}$$
$$\lambda^{c|\pi^*_{z,k}} := \sum_{\delta\in\Delta} \pi^*_{z,k}\Theta_{z,\delta,z'}\mathcal{I}^c_{z'}$$
$$p^{s|\pi^*_{z,k}}(\delta) := \frac{\pi^*_{z,k}\Theta_{z,\delta,z'}\mathcal{I}^s_{z'}}{\lambda^{s|\pi^*_{z,k}}}$$
$$p^{c|\pi^*_{z,k}}(\delta) := \frac{\pi^*_{z,k}\Theta_{z,\delta,z'}\mathcal{I}^c_{z'}}{\lambda^{c|\pi^*_{z,k}}}.$$
5) Then, applying (1), the JSD between p^s_n and p^c_n when $n \to \infty$ is given by:
$$\lim_{n\to\infty} D_i(p^s_n, p^c_n) = H(\{\lambda^s, \lambda^c\}) + \sum_{z:z \text{ is recurrent}} \sum_{k\in\mathbb{N}} p(z,k)\{-H(\{\lambda^{s|\pi^*_{z,k}}, \lambda^{c|\pi^*_{z,k}}\}) + D_i(p^{s|\pi^*_{z,k}}, p^{c|\pi^*_{z,k}})\}. \quad (2)$$
6) When the set of steady state distributions is unique, then in that case, $k = 1$ and we have: $p(z,k) = p(z)$ in (2) above.

Example 4: We revisit Example 3. Then based on Obs_1, the following computation illustrates the steps of JSD computation at site-1.

1) $\sum_z |z| = 14$ and so $\widetilde{\Theta}$ is a 14×14 matrix and
$\pi^* = [\ 0\ 0\ 0\ 0\ 0\ 0\ 0.05\ 0.05\ 0\ 0\ 0.1\ 0.1\ 0.35\ 0.35]$.
Therefore, $p(z_0) = p(z_1) = p(z_2) = 0$, $p(z_3) = 0.1$, $p(z_4) = 0$, and $p(z_5) = 0.9$.
2) Here
$\mathcal{I}^s = [\ 1\ 1\ 1\ 0\ 1\ 0\ 1\ 1\ 1\ 0\ 1\ 1\ 0\ 0\ ]^T$
$\mathcal{I}^c = [\ 0\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ 1\ 1\ ]^T$.
And so, $\lambda^s = 0.3$ and $\lambda^c = 0.7$.
3) Here z_3, and z_5 are recurrent nodes, and each of them forms a SCC. We have $\pi^*_{z_3} = [0.5\ 0.5]$, and while there are multiple solutions to the equation set $\pi^*_{z_5} = \frac{\pi^*_{z_5}\Theta_{z_5,a',z_5}}{||\pi^*_{z_5}\Theta_{z_5,a',z_5}||}$, only $\pi^*_{z_5} = [0.11\ 0.11\ 0.39\ 0.39]$ is reachable. Thus, each set of recurrent nodes is a singleton set, and each with a unique fixed-point distribution. Therefore, for each recurrent node z, $p(z,k) = p(z)$.
4) Here $\mathcal{I}^s_{z_3} = [1\ 1]^T$, $\mathcal{I}^c_{z_3} = [0\ 0]^T$, $\mathcal{I}^s_{z_5} = [1\ 1\ 0\ 0]^T$, $\mathcal{I}^c_{z_5} = [0\ 0\ 1\ 1]^T$. For z_3 and $\pi^*_{z_3}$,
$$\lambda^{s|\pi^*_{z_3}} = 1,\ \lambda^{c|\pi^*_{z_3}} = 0$$
$$p^{s|\pi^*_{z_3}}(a') = \frac{\pi^*_{z_3}\Theta_{z_3,a',z_3}\mathcal{I}^s_{z_3}}{\lambda^{s|\pi^*_{z_3}}} = 1$$
$$p^{s|\pi^*_{z_3}}(b') = p^{c|\pi^*_{z_3}}(b') = p^{c|\pi^*_{z_3}}(a') = 0.$$
For z_5 and $\pi^*_{z_5}$,
$$\lambda^{s|\pi^*_{z_5}} = 0.22,\ \lambda^{c|\pi^*_{z_5}} = 0.78$$

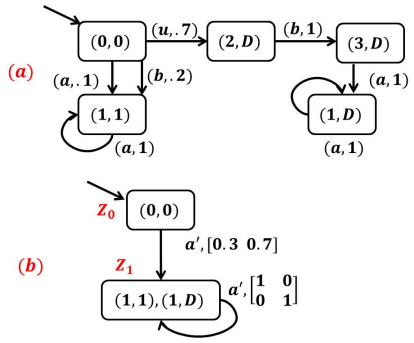

Fig. 6. Model G^R for system of Fig. 2(a) under no collusion; (b) Observer under no collusion.

$$p^{s|\pi^*_{z_5}}(a') = \frac{\pi^*_{z_5}\Theta_{z_5,a',z_5}\mathcal{I}^s_{z_5}}{\lambda^{s|\pi^*_{z_5}}} = 1$$
$$p^{s|\pi^*_{z_5}}(b') = p^{c|\pi^*_{z_5}}(b') = 0$$
$$p^{c|\pi^*_{z_5}}(a') = \frac{\pi^*_{z_5}\Theta_{z_5,a',z_5}\mathcal{I}^c_{z_5}}{\lambda^{c|\pi^*_{z_5}}} = 1.$$

5) Then, we have
$$\lim_{n\to\infty} D_1(p^s_n, p^c_n) = H(\{\lambda^s, \lambda^c\}) + \sum_{z:p(z)>0} p(z)$$
$$\{-H(\{\lambda^{s|\pi^*_z}, \lambda^{c|\pi^*_z}\}) + D_1(p^{s|\pi^*_z}, p^{c|\pi^*_z})\} = \mathbf{0.197}.$$

In contrast, when there is no collusion among observers (so there is no communication among the two sites), Fig. 6(a) and Fig. 6(b) show, respectively, the refined plant G^R (no incoming channels and so identical refined model at all sites) and the corresponding site-1 observer structure. The JSD value, computed in same manner as above but with respect to the observer structure of Fig. 6(b), is simply **Zero**, i.e., no amount of secrets is revealed under no collusion. This is because for every observation, the probability of it coming from secrets in K vs from covers in $L-K$ is exactly the same.

IV. Conclusion

In this paper we studied the problem of secrecy loss quantification in partially-observed discrete event systems (PODESs) in the presence of distributed collusive attackers/observers. The information about system secrets is revealed additionally through the side-channel of observations being exchanged among local observers over bounded delay communications. We proposed a method to compute the secrecy loss in this distributed collusive setting, by introducing bounded-delay channel models as in [10] to extend the system model to capture the effect of exchange of observations, and employing the JSD computation from our earlier work [1] on the extended model to arrive at the measure for secrecy loss. Examples were provided to illustrate our approach. Future work will involve developing a software tool for JSD computation, and performing application studies. Knowing the JSD value can help an engineer to perform secrecy analysis of a system, and revisit its design to make it improve its level of secrecy if needed.

Appendix

We describe the computation of an observer transition structure to track the evolution of $\mathcal{G}^R_i$ over only its observed symbols $\Delta := \cup_{i\in I}\Delta_i$, and the associated transition matrices $\{\Theta(\delta) \mid \delta \in \Delta\}$. Given the refined extended plant model at a site-i, $\mathcal{G}^R_i$, and its extended observation mask $\mathcal{M}_i : \overline{\Sigma} \cup_{j\in I-\{i\}} \overline{\Delta}_j \to \overline{\Delta} = \cup_{i\in I}\overline{\Delta}_i$, define the set of traces originating at $(x,\vec{q},\overline{y})) \in X \times (\Pi_{j\neq i}Q^{(d)}_{ji}) \times \overline{Y}$, terminating at $(x',\vec{q'},\overline{y}')) \in X \times (\Pi_{j\neq i}Q^{(d)}_{ji}) \times \overline{Y}$ and executing a sequence of unobservable events followed by a single observable event with observation $\delta \in \Delta$ as: $L_{\mathcal{G}^R_i}((x,\vec{q},\overline{y}),\delta,(x',\vec{q},\overline{y}')) := \{s \in (\overline{\Sigma} \cup_{j\in I-\{i\}} \overline{\Delta}_j)^* | s = u\sigma, \mathcal{M}_i(u) = \epsilon, \mathcal{M}_i(\sigma) = \delta,\ \gamma((x,\vec{q},\overline{y}),s,(x',\vec{q'},\overline{y}')) > 0\}$. Define its probability, $\alpha(L_{\mathcal{G}^R_i}((x,\vec{q},\overline{y}),\delta,(x',\vec{q'},\overline{y}'))) := \sum_{s\in L_{\mathcal{G}^R_i}((x,\vec{q},\overline{y}),\delta,(x',\vec{q'},\overline{y}'))} \gamma((x,\vec{q},\overline{y}),s,(x',\vec{q'},\overline{y}'))$, and denote it as $\theta_{(x,\vec{q},\overline{y}),\delta,(x',\vec{q'},\overline{y}')}$. Also define $\lambda_{(x,\vec{q},\overline{y}),(x',\vec{q'},\overline{y}')} = \sum_{\sigma\in\Sigma_{uo}} \gamma((x,\vec{q},\overline{y}),\sigma,(x',\vec{q'},\overline{y}'))$ as the probability of transitioning from $(x,\vec{q},\overline{y})$ to $(x',\vec{q'},\overline{y}')$ while executing a single unobservable event. Then, letting $i = (x,\vec{q},\overline{y})$ and $j = (x',\vec{q'},\overline{y}')$, $\theta_{i,\delta,j} = \sum_k \lambda_{i,k}\theta_{k,\delta,j} + \sum_{\sigma\in\overline{\Sigma}\cup_{j\in I-\{i\}}\overline{\Delta}_j:\mathcal{M}_i(\sigma)=\delta} \gamma(i,\sigma,j)$, where the first term on the right hand side (RHS) corresponds to transitioning in at least two steps (i to intermediate k unobservably, and k to j with a single observation δ at the end), whereas, the second term on RHS corresponds to transitioning in exactly one step [15]. Thus, for each $\delta \in \Delta$, all the probabilities $\{\theta_{i,\delta,j} | i,j \in X \times (\Pi_{j\neq i}Q^{(d)}_{ji}) \times \overline{Y}\}$ can be found by solving the following matrix equation [17]: $\Theta(\delta) = \Lambda\Theta(\delta) + \Gamma(\delta)$, where $\Theta(\delta), \Lambda$ and $\Gamma(\delta)$ are all $|X \times (\Pi_{j\neq i}Q^{(d)}_{ji}) \times \overline{Y}| \times |X \times (\Pi_{j\neq i}Q^{(d)}_{ji}) \times \overline{Y}|$ square matrices whose ijth elements are given by $\theta_{i,\delta,j}, \lambda_{i,j}$ and $\sum_{\sigma\in\overline{\Sigma}\cup_{j\in I-\{i\}}\overline{\Delta}_j:\mathcal{M}_i(\sigma)=\delta} \gamma(i,\sigma,j)$, respectively.

References

[1] J. Chen, M. Ibrahim, and R. Kumar, "Quantification of secrecy in partially observed stochastic discrete event systems," *IEEE Transactions on Automation Science and Engineering*, accepted (Sept. 2015).

[2] G. Smith, "On the foundations of quantitative information flow," in *Proc. Int. Conf. Foundations of Software Science and Computation Structures (FoSSaCS 09)*, 2009, pp. 288–302.

[3] B. Espinoza and G. Smith, "Min-entropy as a resource," *Information and Computation*, vol. 226, pp. 57–75, Apr. 2013.

[4] S. Takai and R. Kumar, "Verification and synthesis for secrecy in discrete-event systems," in *Proc. IEEE American Control Conference, (ACC '09)*, St. Louis, MO, Jun. 2009, pp. 4741–4746.

[5] J. Bryans, M. Koutny, and C. Mu, "Towards quantitative analysis of opacity," *Technical Reports Series, Newcastle University*, Nov. 2011.

[6] A. Saboori and C. N. Hadjicostis, "Probabilistic current-state opacity is undecidable," in *Proc. of 19th Int. Symp. Math. Theory Netw. and Syst. (MTNS '2010)*, Budapest, Hungary, Jul. 2010, pp. 477–483.

[7] A. Saboori and C. N. Hadjicostis, "Opacity verification in stochastic discrete event systems," in *Proc. 49th IEEE Conference on Decision and Control*, Atlanta, GA, Dec. 2010, pp. 6759–6764.

[8] M. Ibrahim, J. Chen, and R. Kumar, "Secrecy in stochastic discrete event systems," in *Proc. of 11th IEEE International Conference on Networking, Sensing and Control (ICNSC'14)*, Miami, FL, Apr. 2014, pp. 48–53.

[9] R. Jacob, J.-J. Lesage, and J.-M. Faure, "Opacity of discrete event systems: models, validation and quantification," in *Proc. the 5th international workshop on Dependable Control of Discrete Systems (DCDS'15), hal-01139890*, Cancun, Mexico, May 2015.

[10] W. Qiu and R. Kumar, "Distributed diagnosis under bounded-delay communication of immediately forwarded local observations," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 38, pp. 628–643, May 2008.

[11] V. K. Garg, R. Kumar, and S. I. Marcus, "A probabilistic language formalism for stochastic discrete-event systems," *IEEE Transactions on Automatic Control*, vol. 44, no. 2, pp. 280–293, Feb. 1999.

[12] A. Xie and P. A. Beerel, "Efficient state classification of finite-state Markov chains," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 17, no. 12, pp. 1334–1339, Dec. 1998.

[13] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley and Sons, 2012.

[14] J. Lin, "Divergence measures based on the shannon entropy," *IEEE Trans. on Information Theory*, vol. 37, no. 1, pp. 145–151, Jan. 1991.

[15] J. Chen and R. Kumar, "Failure detection framework for stochastic discrete event systems with guaranteed error bounds," *IEEE Trans. on Automatic Control*, vol. 60, no. 6, pp. 1542–1553, Jun. 2015.

[16] T. Kaijser, "A limit theorem for partially observed markov chains," *The Annals of Probability*, vol. 3, no. 4, pp. 677–696, Aug. 1975.

[17] X. Wang and A. Ray, "A language measure for performance evaluation of discrete-event supervisory control systems," *Applied Math. Modelling*, vol. 28, no. 9, pp. 817–833, Sep. 2004.