An Information Theoretic Measure for Secrecy Loss in Stochastic Discrete Event Systems

Mariam Ibrahim[1&2], 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,junchen,rkumar}@iastate.edu

***Abstract*—While cryptography is used to protect the content of secret information (message) by making it undecipherable, behaviors (as opposed to information) may not be encrypted, and may only be protected by partially or fully hiding through creation of ambiguity by providing covers that generate indistinguishable observations from secrets. Having a cover together with partial observability does cause ambiguity about the system behaviors to be kept secret, yet some information about secrets may still be leaked due to statistical difference between the occurrence probabilities of the secrets and their covers. One possible quantification of statistical difference between two distributions is based on their Jenson-Shannon divergence (JSD). We propose a computation of JSD for systems modeled as partially-observed Markov chains (POMC). Since an adversary is likely to discriminate more if he/she observes for a longer period, our goal is to evaluate the worst-case loss of secrecy as obtained in limit over longer and longer observations. Illustrative example is provided to demonstrate the proposed computation approach.**

Keywords*—*Partially-observed Markov chains (POMC), Jenson-Shannon divergence (JSD), Secrecy quantification.

I. Introduction

The rapid progress in information and communication technology has made it possible for an adversary to eavesdrop and/or attack confidential or private communication. While cryptography is used to protect the content of secret information (message) by making it undecipherable, the same technique may not be used to hide behaviors which may not be encrypted. In such cases, *secrecy* can instead be attained through creation of ambiguity, caused for example by partial observation that ambiguates secrets from covers. Researchers in the field of security and privacy have explored many techniques for hiding secrets based on ambiguation schemes such as, *Steganography and Watermarking* [1], [2], *Network level Anonymization* [3], and *Software Obfuscation* [4].

Various notions of information secrecy have been explored in literature. References [5], [6], [7] defined the non-interference for input-output systems as a property in which the outputs that are observable to an *adversary* should not depend on any *secret* input so that the adversary does not deduce anything about the secret input by observing the output. Non-interference is a logical notion that is either satisfied or violated, and as such it does not allow the quantification of the degree to which a system may violate the property. To circumvent this limitation, the notion is enriched for probabilistic systems for which the amount of interference can be quantified in terms of the mutual information between the inputs and the outputs [5]. This however is only an average case measure, and a worst case measure can be obtained by replacing entropy with *min-entropy* in the definition of mutual information [7]. For secrecy over sequences of inputs/outputs, [8] requires that the execution of behaviors constituting a secret must not be revealed to an observer by masking those behaviors through indistinguishable behaviors that are non-secret, known as *cover*. This is indeed analogous to the notion of non-interference, which by virtue of being logical has the same limitation that it cannot quantify the degree to which a system is interfering (or leaks information).

For probabilistic DESs, where each discrete transition is associated with a certain occurrence probability, more powerful notions of secrecy can be defined. For example, [9] used Jensen-Shannon divergence between the distributions of a secret versus its cover as a way to quantify the secrecy. The computation of Jensen-Shannon divergence is not known in general: Only an approximation algorithm for upper bounding the values of JSD was provided in [9]. Another attempt to generalize secrecy from logical to stochastic DESs is provided in [10], where, alike the setting of mutual information based characterization of information leakage, the authors consider the difference between the prior and posterior distributions (before and after any observations) of the secret states, and require it to be upper bounded. The corresponding verification problem turns out to be undecidable. In another paper [11], the same authors proposed the notion of *Step-Based Almost Current-State Opacity* requiring the probability of revealing the secret must be upper bounded at each time step. This notion is decidable, but stringent since it is defined for each individual step. In contrast, another definition of S_τ-secrecy proposed by us [12], bounds the probability of revealing the secret over the set of *all* behaviors, as opposed to for each step. We showed that S_τ-Secrecy can be viewed as a generalization of the logical secrecy defined in [8], and that it is a variant of the divergence used in [9]. The above mentioned works on secrecy (also referred to *opacity* in literatures), along with related articles have been reviewed in a recent survey [13]. The work reported in [14] also uses JSD measure for determining statistical difference in Markovian models of genetic sequences from phylogenetically proximal organisms, which however is not related to secrecy as no information hiding through partial observation is involved.

In this paper, we propose a JSD based quantification to measure the secrecy loss in stochastic discrete event systems. Different from the above mentioned works, we consider Markovian generators of *arbitrary* long sequences and that are *partially-observed*, and provide a *recursive* method for JSD computation: Given the distribution with respect to length-$(n-1)$ sequences, and the length-1 dynamics of the underlying

This research was supported in part by PNNL and John Deere through NSF-IUCRC, Security and Software Engineering Research Center (S2ERC), and the National Science Foundation under the grants NSF-ECCS-0926029 and NSF-CCF-1331390.

partially-observed model, it computes the JSD of length-(n) sequences. Under certain conditions, this recursion reaches a fixed point, measuring worst case statistical difference that is defined over arbitrary long sequences. Since JSD is always bounded between 1 and 0, this worst case value is also bounded. In this paper, we derive the above recursion, and next construct an observer model of the given POMC, which we then use to develop a state-based computation of the fixed point JSD measure. The computation of JSD for a POMC is challenging since a finite-state Markov chain under partial observations is potentially infinite-state (with the state-space being the conditional state distributions following the observations). However, a finite-state observer representation is possible, which we construct and employ for divergence computation. This observer model is not a Markov chain model since the transition probabilities are no longer scalars, rather matrices, not necessarily square.

Rest of the paper is organized as follows. Section II presents notation and preliminaries. Divergence based secrecy quantification of information-flow secrecy is presented in Section III, whereas Section IV presents an observer based computation of worst-case JSD measure resulting from arbitrary long observations. Section V presents example to illustrate the approach, while Section VI concludes the paper.

II. Notations and Preliminaries

A. Stochastic DESs

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^*, and $\Sigma^+ := \Sigma^* - \{\epsilon\}$. 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 DES can be modeled by a *stochastic automaton* $G = (X, \Sigma, \alpha, x_0)$ that is an initialized labeled Markov chain, 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 [15], and $\forall x \in X, \sum_{\sigma \in \Sigma} \sum_{x' \in X} \alpha(x, \sigma, x') = 1$. G is non-stochastic if $\alpha : X \times \Sigma \times X \to \{0, 1\}$, and a non-stochastic DES is deterministic if $\forall x \in X, \sigma \in \Sigma, \sum_{x' \in X} \alpha(x, \sigma, x') \leq 1$. The transition probability function α can be generalized to $\alpha : X \times \Sigma^* \times X$ in a natural way. 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 is defined. 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 same size as number of states, 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 fixed-point of $\pi^* = \pi^* \Omega^d$). Then any state i is recurrent if and only if the ith entry of π^* is nonzero. Identifying the set of recurrent states can be done polynomially, by the algorithm presented in [16].

The events executed by a DES can be partially observed by an observer (i.e., an adversary). The limited observation capability of an observer can be represented as an observation mask, $M : \overline{\Sigma} \to \overline{\Delta}$, where Δ is the set of observed symbols and $M(\epsilon) = \epsilon$. An event σ is unobservable if $M(\sigma) = \epsilon$. The set of unobservable events is denoted as Σ_{uo} and the set of observable events is then given by $\Sigma - \Sigma_{uo}$. The observation mask can be generalized in natural way to Σ^* with $M(\epsilon) = \epsilon$ and $\forall s \in \Sigma^*, \sigma \in \overline{\Sigma}, M(s\sigma) = M(s)M(\sigma)$.

B. Secret/non-secret behaviors and refined plant

Suppose $K \subseteq \Sigma^*$ models the secret behaviors (traces), whereas the remaining traces in $L - K$ can be viewed as its cover. Let the stochastic automaton $G = (X, \Sigma, \alpha, x_0)$ with generated language $L(G) = L$ be the system model, and the *deterministic* automaton $R = (Y, \Sigma, \beta, y_0)$ which specifies the secret behaviors K be such that $L(R) = K$. Then a refinement of G with respect to R, denoted G^R, can be used to capture the property-satisfying/violating traces in form of the reachability of certain non-secret states (the state has D in it's second coordinate), and is given by $G^R := (X \times \overline{Y}, \Sigma, \gamma, (x_0, y_0))$, where $\overline{Y} = Y \cup \{D\}$, and $\forall (x, \overline{y}), (x', \overline{y}') \in X \times \overline{Y}, \sigma \in \Sigma, \gamma((x, \overline{y}), \sigma, (x', \overline{y}')) = \alpha(x, \sigma, x')$ 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 $\gamma((x, \overline{y}), \sigma, (x', \overline{y}')) = 0$. Then it can be seen that the refined plant G^R has the following properties: (1) $L(G^R) = L(G)$; (2) any property-satisfying trace $s \in L(G)$ but not in $L(R)$ transitions the refinement G^R to a non-secret state; (3) for each $s \in L(G) = L(G^R)$, $\sum_{x \in X} \alpha(x_0, s, x) = \sum_{(x, \overline{y}) \in X \times \overline{Y}} \gamma((x_0, y_0), s, (x, \overline{y}))$, i.e., the occurrence probability of each trace in G^R is the same as that in G. For $(x, \overline{y}), (x', \overline{y}') \in X \times \overline{Y}$, and $\delta \in \Delta$, define the set of traces originating at $(x, \overline{y})$, terminating at $(x', \overline{y}')$ and executing a sequence of unobservable events followed by a single observable event with observation δ as $L_{G^R}((x, \overline{y}), \delta, (x', \overline{y}')) := s \in \Sigma^* | s = u\sigma, M(u) = \epsilon, M(\sigma) = \delta, \gamma((x, \overline{y}), s, (x', \overline{y}')) > 0$. Define $\alpha(L_{G^R}((x, \overline{y}), \delta, (x', \overline{y}'))) := \sum_{s \in L_{G^R}((x, \overline{y}), \delta, (x', \overline{y}'))} \gamma((x, \overline{y}), s, (x', \overline{y}'))$, and denote it as $\theta_{i,\delta,j}$, i.e., it is the probability of all traces originating at $i = (x, \overline{y})$, terminating at $j = (x', \overline{y}')$ and executing a sequence of unobservable events followed by a single observable event with observation δ. Also define $\lambda_{ij} = \sum_{\sigma \in \Sigma_{uo}} \gamma((x, \overline{y}), \sigma, (x', \overline{y}'))$ as the probability of transitioning from $(x, \overline{y})$ to $(x', \overline{y}')$ while executing a single unobservable event. Then $\theta_{i,\delta,j} = \sum_k \lambda_{ik} \theta_{k,\delta,j} + \sum_{\sigma \in \Sigma : M(\sigma) = \delta} \gamma((x, \overline{y}), \sigma, (x', \overline{y}'))$, where the first term on the right hand side (RHS) corresponds to transitioning in at least two steps (i to intermediate k

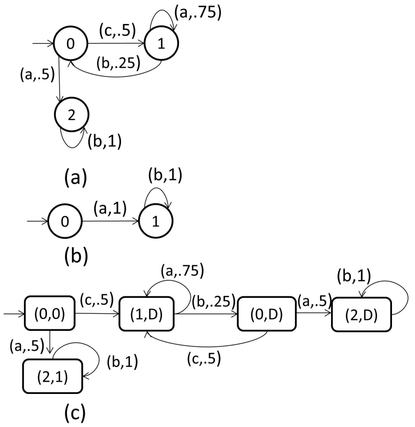

Fig. 1. (a) Stochastic automaton G; (b) deterministic secret specification R; (c) refinement G^R;

unobservaby, 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 [17]. Thus, for each $\delta \in \Delta$, all the probabilities $\{\theta_{i,\delta,j} | i,j \in X \times \overline{Y}\}$ can be found by solving the following matrix equation [18]:

$$\Theta(\delta) = \Lambda\Theta(\delta) + \Gamma(\delta), \tag{1}$$

where $\Theta(\delta), \Lambda$ and $\Gamma(\delta)$ are all $|X \times \overline{Y}| \times |X \times \overline{Y}|$ square matrices whose ijth elements are given by $\theta_{i,\delta,j}, \lambda_{ij}$ and $\sum_{\sigma\in\Sigma:M(\sigma)=\delta} \gamma((x,\overline{y}),\sigma,(x',\overline{y}'))$, respectively.

Example 1: Fig. 1(a) is an example of a stochastic automaton G. The set of states is $X = \{0,1,2\}$ with initial state $x_0 = 0$, event set $\Sigma = \{a,b,c\}$. A state is depicted as a node, whereas a transition is depicted as an edge between its origin and termination states, with its event name and probability value labeled on the edge. The observation mask M is such that $M(c) = \epsilon$ and for all other events $\sigma \in \{a,b\}$, $M(\sigma) = \sigma$. Suppose R is given in Fig. 1(b), i.e., $K = L(R) = ab^*, L - K = ca^* \cup (ca^*b)^+ \cup (ca^*b)^+ab^*$. Then the refinement G^R automaton is shown in Fig. 1(c). Let the state space of G^R be $Y = y_1 = (0,0), y_2 = (2,1), y_3 = (1,D), y_4 = (0,D), Y_5 = (2,D)$. Then, by solving (1) we get for $\Theta(a)$ the following entries: $\Theta(1,2) = \Theta(4,5) = 0.5, \Theta(1,3) = \Theta(4,3) = 0.375, \Theta(3,3) = 0.75$, and zeros elsewhere. Similarly, we can solve for $\Theta(b)$ entries as follows: $\Theta(1,4) = \Theta(4,4) = 0.125, \Theta(2,2) = \Theta(5,5) = 1, \Theta(3,4) = 0.25$, and zeros elsewhere. Note that the size of $|\Theta(a)| = |\Theta(b)| = 5 \times 5$ matrices.

In [8] a logical version of secrecy was defined, which is satisfied whenever each secret can be masked by a cover, and vice-versa, with non-zero probability. A weaker version considered in [12], allows some secrets/covers to be non-masked, but limits the probability of such traces to be a small number. In the next section a new approach for measuring the level of secrecy is introduced utilizing the notion of JSD.

III. Divergence based Secrecy Quantification

For any $n \in \mathbb{N}$, and a length-n observation $o \in \Delta^n$, let $p_n(o)$ denote the probability of observation o. 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:

$$\begin{aligned} H(p_n) &= -\sum_{o\in\Delta^n} p_n(o)\log p_n(o) \\ &= H(p_{n-1}) - \sum_{o\in\Delta^{n-1}} p_{n-1}(o) \sum_{\delta\in\Delta} p(\delta|o)\log p(\delta|o) \end{aligned} \tag{2}$$

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 $p_n^c(o)$:

$$\begin{aligned} p_n^s(o) &:= \frac{Pr(s \in K \cap M^{-1}(o))}{Pr(s \in K \cap M^{-1}(\Delta^n))} =: \frac{\widetilde{p}_n^s(o)}{\lambda_n^s}, \\ p_n^c(o) &:= \frac{Pr(s \in (L-K) \cap M^{-1}(o))}{Pr(s \in (L-K) \cap M^{-1}(\Delta^n))} =: \frac{\widetilde{p}_n^c(o)}{\lambda_n^c}. \end{aligned}$$

Note λ_n^s and λ_n^c are the probabilities of secrets and covers, respectively, of length n, and $\lambda_n^s + \lambda_n^c = 1$ for all $n \in \mathbb{N}$. Then the entropy of p_n^s and p_n^c are given, respectively, by:

$$H(p_n^s) = -\sum_{o\in\Delta^n} p_n^s(o)\log p_n^s(o) \tag{3}$$

$$H(p_n^c) = -\sum_{o\in\Delta^n} p_n^c(o)\log p_n^c(o) \tag{4}$$

The ability of an intruder 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, denoted $D(p_n^s, p_n^c)$. This JSD is given by the following weighted sum of a pair of KL-divergences between, respectively, p_n^s and p_n^c, and their weighted sum:

$$\begin{aligned} D(p_n^s, p_n^c) &= \lambda_n^s D_{KL}(p_n^s, \lambda_n^s p_n^s + \lambda_n^c p_n^c) \\ &\quad + \lambda_n^c D_{KL}(p_n^c, \lambda_n^s p_n^s + \lambda_n^c p_n^c) \\ &= \lambda_n^s \sum_{o\in\Delta^n} p_n^s(o)\log\frac{p_n^s(o)}{\lambda_n^s p_n^s(o) + \lambda_n^c p_n^c(o)} \\ &\quad + \lambda_n^c \sum_{o\in\Delta^n} p_n^c(o)\log\frac{p_n^c(o)}{\lambda_n^s p_n^s(o) + \lambda_n^c p_n^c(o)} \\ &= H[\lambda_n^s p_n^s + \lambda_n^c p_n^c] - \lambda_n^s H(p_n^s) - \lambda_n^c H(p_n^c). \end{aligned} \tag{5}$$

where D_{KL} represents the Kullback-Leibler (KL) divergence. Note that JSD is symmetric in its arguments and bounded by 0 and 1. 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: $\lim_{n\to\infty} D(p_n^s, p_n^c)$. This worst-case JSD provides an upper bound to quantification of the amount of information leaked about secrets.

A. Recursive Characterization

We first develop a recursive computation for $D(p_n^s, p_n^c)$, relating it to distributions of length-$(n-1)$ observations, and divergence of length-1 distributions. For $o \in \Delta^*$ and $\delta \in \Delta$, define the distributions of secret versus cover upon a single observation δ following a history of observation o:

$$\begin{aligned} p^s(\delta|o) &:= \frac{Pr(s \in K \cap M^{-1}(o\delta))}{Pr(s \in K \cap M^{-1}(o\{\Delta\}))} \\ &=: \frac{\widetilde{p}^s(\delta|o)}{\sum_{\delta\in\Delta} \widetilde{p}^s(\delta|o)} = \frac{\widetilde{p}^s(\delta|o)}{\lambda^{s|o}}, \text{and} \qquad (6) \\ p^c(\delta|o) &:= \frac{Pr(s \in (L-K) \cap M^{-1}(o\delta))}{Pr(s \in (L-K) \cap M^{-1}(o\{\Delta\}))} \\ &=: \frac{\widetilde{p}^c(\delta|o)}{\sum_{\delta\in\Delta} \widetilde{p}^c(\delta|o)} = \frac{\widetilde{p}^c(\delta|o)}{\lambda^{c|o}}. \qquad (7) \end{aligned}$$

Then note also that $\lambda^{c|o} + \lambda^{s|o} = 1$, and the JSD of $p^s(\cdot|o)$ and $p^c(\cdot|o)$, denoted for short as $p^{s|o}$ and $p^{c|o}$ respectively, satisfies:

$$\begin{aligned} D(p^{s|o}, p^{c|o}) &= H[\lambda^{s|o} p^{s|o} + \lambda^{c|o} p^{c|o}] - \lambda^{s|o} H(p^{s|o}) \\ &\quad -\lambda^{c|o} H(p^{c|o}) \qquad (8) \\ &= -\sum_{\delta\in\Delta} p(\delta|o) \log p(\delta|o) + \sum_{\delta\in\Delta} \widetilde{p}^s(\delta|o) \log p^s(\delta|o) \\ &\quad + \sum_{\delta\in\Delta} \widetilde{p}^c(\delta|o) \log p^c(\delta|o), \qquad (9) \end{aligned}$$

where $p(\delta|o) := \lambda^{s|o} p^{s|o} + \lambda^{c|o} p^{c|o} = \widetilde{p}^s(\delta|o) + \widetilde{p}^c(\delta|o)$. By substituting $p^s(\delta|o)$ in (9) with $\frac{\widetilde{p}^s(\delta|o)}{\lambda^{s|o}}$ as in (6) and $p^c(\delta|o)$ in (9) with $\frac{\widetilde{p}^c(\delta|o)}{\lambda^{c|o}}$ as in (7), we have

$$\begin{aligned} D(p^{s|o}, p^{c|o}) &= -\sum_{\delta\in\Delta} p(\delta|o) \log p(\delta|o) + H(\{\lambda^{s|o}, \lambda^{c|o}\}) \\ &\quad + \sum_{\delta\in\Delta} \widetilde{p}^s(\delta|o) \log \widetilde{p}^s(\delta|o) + \sum_{\delta\in\Delta} \widetilde{p}^c(\delta|o) \log \widetilde{p}^c(\delta|o). \end{aligned}$$

Then

$$\begin{aligned} D(p^{s|o}, p^{c|o}) - H(\{\lambda^{s|o}, \lambda^{c|o}\}) &= -\sum_{\delta\in\Delta} p(\delta|o) \log p(\delta|o) \\ &\quad + \sum_{\delta\in\Delta} \widetilde{p}^s(\delta|o) \log \widetilde{p}^s(\delta|o) + \sum_{\delta\in\Delta} \widetilde{p}^c(\delta|o) \log \widetilde{p}^c(\delta|o) \end{aligned}$$

Then we have

$$\begin{aligned} D(p^s_n, p^c_n) &= H(\{\lambda^s_n, \lambda^c_n\}) \\ &\quad + \sum_{o\in\Delta^{n-1}} p_{n-1}(o)\{-H(\{\lambda^{s|o}, \lambda^{c|o}\}) + D(p^{s|o}, p^{c|o})\} \qquad (10) \end{aligned}$$

B. State Distribution based characterization

We have characterized the JSD computation following observations of length n, and next we map it to a computation based on the state-distribution following an 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 $\pi(o\delta) = \frac{\pi(o)\times\Theta(\delta)}{||\pi(o)\times\Theta(\delta)||}$, where π_0 is the initial state distribution. 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. Then the divergence of (10) can be rewritten as:

$$\begin{aligned} D(p^s_n, p^c_n) &= H(\{\lambda^s_n, \lambda^c_n\}) \\ &\quad + \sum_{\pi\in\Pi} P_{n-1}(\pi)\{-H(\{\lambda^{s|\pi}, \lambda^{c|\pi}\}) + D(p^{s|\pi}, p^{c|\pi})\}, \qquad (11) \end{aligned}$$

where for each $\pi \in \Pi$, $p^{s|\pi} = p^s(\cdot|\pi), p^{c|\pi} = p^c(\cdot|\pi), \lambda^{s|\pi}, \lambda^{c|\pi}$ are defined as follows, in which the notations $\mathcal{I}^s$ 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 $L-K$):

$$\begin{aligned} \widetilde{p}^s(\delta|\pi) &:= \pi\Theta(\delta)\mathcal{I}^s, \widetilde{p}^c(\delta|\pi) := \pi\Theta(\delta)\mathcal{I}^c \\ \lambda^{s|\pi} &:= \sum_{\delta\in\Delta} \widetilde{p}^s(\delta|\pi), \lambda^{c|\pi} := \sum_{\delta\in\Delta} \widetilde{p}^c(\delta|\pi) \\ p^s(\delta|\pi) &:= \frac{\widetilde{p}^s(\delta|\pi)}{\lambda^{s|\pi}}, p^c(\delta|\pi) := \frac{\widetilde{p}^c(\delta|\pi)}{\lambda^{c|\pi}}. \end{aligned}$$

In the limit when $n \to \infty$, if the distribution $P_n(\cdot)$ over Π converges to $P^*(\cdot)$, then the limit of $D(p^s_n, p^c_n)$ exists (see for example [19] for a condition under which such a convergence is guaranteed).

IV. Observer based computation

Let Obs be an observer automaton with state set $Z \subseteq 2^{X\times\overline{Y}}$, so that each node $z \in Z$ of the observer is a subset of the system states, i.e., $z \subseteq (X, \overline{Y})$, and we use $|z|$ to denote the number of system states in z. Obs is initialized at node $z_0 = \{(x_0, y_0)\}$, 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., $z' = \{(x', \overline{y}') \in X \times \overline{Y} : \exists (x, \overline{y}) \in z, L_{G^R}((x,\overline{y}), \delta, (x', \overline{y}')) \neq \emptyset\}$. Associated with this transition is the transition probability matrix $\Theta_{z,\delta,z'}$ of size $|z|$ by $|z'|$, and a submatrix of Θ matrix introduced earlier, whose ijth element is $\theta_{i,\delta,j}$, which is the transition probability from ith element $(x, \overline{y})$ of z to jth element $(x', \overline{y}')$ of z' while producing the observation δ, and equals $\alpha(L_{G^R}((x,\overline{y}), \delta, (x', \overline{y}')))$.

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 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. 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 following o and $o\delta$ respectively. Then it can be seen that along any cycle in Obs, 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. In the following, we assume the existence of such steady-state:

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

The following procedure computes the worst-case loss of secrecy, under Assumption 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 $\{z_1, z_2, \dots, z_n\}$ that form a SCC, define a set of distributions $\{\pi^*_{z_1}, \pi^*_{z_2}, \dots, \pi^*_{z_n}\}$ to be a set of steady state distributions if the following holds:

$$\forall i, j, \delta, \text{ for which } \Theta_{z_i,\delta,z_j} \text{ is defined,}$$
$$\pi^*_{z_j} = \frac{\pi^*_{z_i} \Theta_{z_i,\delta,z_j}}{||\pi^*_{z_i} \Theta_{z_i,\delta,z_j}||}. \quad (12)$$

Then $\pi^*_{z_i}$ represents a steady state conditional distribution following a single sufficiently long observation, that reaches z_i. Due to the fixed-point nature of $\pi^*_{z_i}$, any other longer observation that also reaches z_i also induces the same conditional distribution $\pi^*_{z_i}$. There may exist multiple set of steady state distributions for a given set of recurrent nodes, denoted say as $\{\{\pi^*_{z_1,k}, \dots, \pi^*_{z_n,k}\}, k \in \mathbb{N}\}$. Then under Assumption 1, 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 $p(z,k) := Pr[\{o \mid o \text{ reaches } z \text{ and } \pi(o) = \pi^*_{z,k}\}]$.

Then the following formulas compute $D(p^s_n, p^c_n)$ as $n \to \infty$, under Assumption 1.
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:

$$\widetilde{p}^s(\delta|\pi^*_{z,k}) := \pi^*_{z,k}\Theta_{z,\delta,z'}\mathcal{I}^s_{z'}, \widetilde{p}^c(\delta|\pi^*_{z,k}) := \pi^*_{z,k}\Theta_{z,\delta,z'}\mathcal{I}^c_{z'}$$
$$\lambda^{s|\pi^*_{z,k}} := \sum_{\delta \in \Delta} \widetilde{p}^s(\delta|\pi^*_{z,k}), \lambda^{c|\pi^*_{z,k}} := \sum_{\delta \in \Delta} \widetilde{p}^c(\delta|\pi^*_{z,k})$$
$$p^s(\delta|\pi^*_{z,k}) := \frac{\widetilde{p}^s(\delta|\pi^*_{z,k})}{\lambda^{s|\pi^*_{z,k}}}, p^c(\delta|\pi^*_{z,k}) := \frac{\widetilde{p}^c(\delta|\pi^*_{z,k})}{\lambda^{c|\pi^*_{z,k}}}$$

5) Then, applying (11), the JSD between p^s_n and p^c_n when $n \to \infty$ is given by:

$$\lim_{n\to\infty} D(p^s_n, p^c_n) = H(\{\lambda^s, \lambda^c\})$$
$$+ \sum_{z: z \text{ is recurrent}} \sum_{k \in \mathbb{N}} p(z,k)\{-H(\{\lambda^{s|\pi^*_{z,k}}, \lambda^{c|\pi^*_{z,k}}\})$$
$$+ D(p^{s|\pi^*_{z,k}}, p^{c|\pi^*_{z,k}})\}, \quad (13)$$

where JSD $D(p^{s|\pi^*_{z,k}}, p^{c|\pi^*_{z,k}})$ of $p^s(\delta|\pi^*_{z,k})$ and $p^c(\delta|\pi^*_{z,k})$ can be computed in a similar way as (8).
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 (13) above.

V. Illustrative Example

Consider the models of Fig.(2). The following computation illustrates the steps given in previous section.

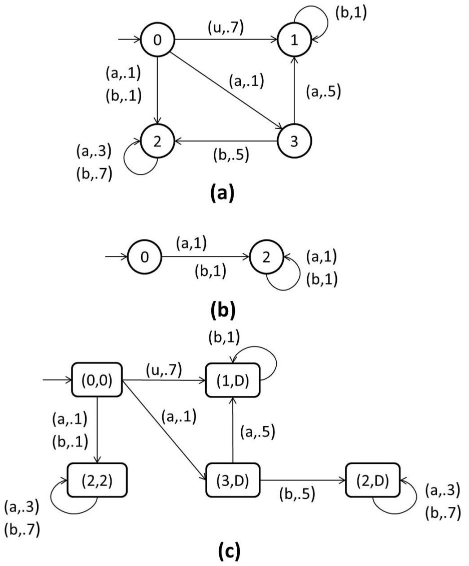

Fig. 2. System model

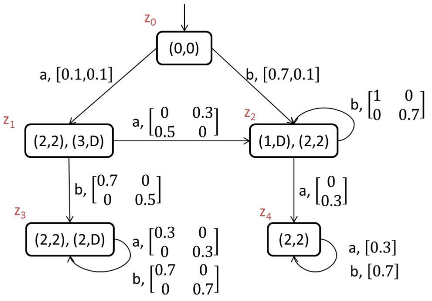

Fig. 3. Observer

1) In this example $\sum_z |z| = 8$ and so $\widetilde{\Theta}$ is a 8×8 matrix with entries: $\widetilde{\Theta}(1,2) = \widetilde{\Theta}(1,3) = \widetilde{\Theta}(1,5) = 0.1$, $\widetilde{\Theta}(3,4) = \widetilde{\Theta}(3,7) = 0.5$, $\widetilde{\Theta}(1,4) = \widetilde{\Theta}(2,6) = \widetilde{\Theta}(5,5) = 0.7$, $\widetilde{\Theta}(2,5) = \widetilde{\Theta}(5,8) = 0.3$, $\widetilde{\Theta}(4,4) = \widetilde{\Theta}(6,6) = \widetilde{\Theta}(7,7) = \widetilde{\Theta}(8,8) = 1$, and zeros elsewhere. and

$$\pi^* = [\begin{array}{cccccccc} 0 & 0 & 0 & 0.75 & 0 & 0.07 & 0.05 & 0.13 \end{array}].$$

Therefore $p(z_0) = p(z_1) = 0$, $p(z_2) = 0.75$, $p(z_3) = 0.12$ and $p(z_4) = 0.13$.
2) Here

$$\mathcal{I}^s = [\begin{array}{cccccccc} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \end{array}]^T$$
$$\mathcal{I}^c = [\begin{array}{cccccccc} 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \end{array}]^T$$

And so $\lambda^s = 0.2$ and $\lambda^c = 0.8$.
3) Here z_2, z_3 and z_4 are recurrent nodes, and each of them forms a SCC. We have $\pi^*_{z_2} = [1\ 0]$, $\pi^*_{z_4} = [1]$, and while there are multiple solutions to the equation set $\pi^*_{z_3} = \frac{\pi^*_{z_3}\Theta_{z_3,a,z_3}}{\pi^*_{z_3}\Theta_{z_3,a,z_3}}$ and $\pi^*_{z_3} = \frac{\pi^*_{z_3}\Theta_{z_3,b,z_3}}{\pi^*_{z_3}\Theta_{z_3,b,z_3}}$, only $\pi^*_{z_3} = [0.5833\ 0.4167]$ is reachable. Thus each recurrent set of nodes is a singleton set, and each with a unique fixed-point distribution. So, for each recurrent node z, $k = 1$ in definition of $p(z,k)$, and hence $p(z,k) = p(z)$.

4) Here $\mathcal{I}^s_{z_2} = [0\ 1]^T$, $\mathcal{I}^c_{z_2} = [1\ 0]^T$, $\mathcal{I}^s_{z_3} = [1\ 0]^T$, $\mathcal{I}^c_{z_3} = [0\ 1]^T$, $\mathcal{I}^s_{z_4} = [1]^T$ and $\mathcal{I}^c_{z_4} = [0]^T$. For z_2 and $\pi^*_{z_2}$,

$$\begin{aligned}
\widetilde{p}^s(a|\pi^*_{z_2}) &= \pi^*_{z_2}\Theta_{z_2,a,z_4}\mathcal{I}^s_{z_4} = 0\\
\widetilde{p}^s(b|\pi^*_{z_2}) &= \pi^*_{z_2}\Theta_{z_2,b,z_2}\mathcal{I}^s_{z_2} = 0\\
\widetilde{p}^c(a|\pi^*_{z_2}) &= \pi^*_{z_2}\Theta_{z_2,a,z_4}\mathcal{I}^c_{z_4} = 0\\
\widetilde{p}^c(b|\pi^*_{z_2}) &= \pi^*_{z_2}\Theta_{z_2,b,z_2}\mathcal{I}^c_{z_2} = 1\\
\lambda^{s|\pi^*_{z_2}} &= \sum_{\delta\in\Delta}\widetilde{p}^s(\delta|\pi^*_{z_2}) = 0\\
\lambda^{c|\pi^*_{z_2}} &= \sum_{\delta\in\Delta}\widetilde{p}^c(\delta|\pi^*_{z_2}) = 1\\
p^c(b|\pi^*_{z_2}) &= \frac{\widetilde{p}^c(b|\pi^*_{z_2})}{\lambda^{c|\pi^*_{z_2}}} = 1\\
p^s(a|\pi^*_{z_2}) &= p^c(a|\pi^*_{z_2}) = p^s(b|\pi^*_{z_2}) = 0
\end{aligned}$$

For z_3 and $\pi^*_{z_3}$,

$$\begin{aligned}
\widetilde{p}^s(a|\pi^*_{z_3}) &= 0.175, \widetilde{p}^s(b|\pi^*_{z_3}) = 0.4083\\
\widetilde{p}^c(a|\pi^*_{z_3}) &= 0.125, \widetilde{p}^c(b|\pi^*_{z_3}) = 0.2917\\
\lambda^{s|\pi^*_{z_3}} &= 0.5833, \lambda^{c|\pi^*_{z_3}} = 0.4167\\
p^s(a|\pi^*_{z_3}) &= 0.3, p^s(b|\pi^*_{z_3}) = 0.7\\
p^c(a|\pi^*_{z_3}) &= 0.3, p^c(b|\pi^*_{z_3}) = 0.7
\end{aligned}$$

For z_4 and $\pi^*_{z_4}$,

$$\begin{aligned}
\widetilde{p}^s(a|\pi^*_{z_4}) &= 0.3, \widetilde{p}^s(b|\pi^*_{z_4}) = 0.7\\
\widetilde{p}^c(a|\pi^*_{z_4}) &= 0, \widetilde{p}^c(b|\pi^*_{z_4}) = 0\\
\lambda^{s|\pi^*_{z_4}} &= 1, \lambda^{c|\pi^*_{z_4}} = 0\\
p^s(a|\pi^*_{z_4}) &= 0.3, p^s(b|\pi^*_{z_4}) = 0.7\\
p^c(a|\pi^*_{z_4}) &= p^c(b|\pi^*_{z_4}) = 0
\end{aligned}$$

5) Therefore we have,

$$\begin{aligned}
\lim_{n\to\infty} D(p^s_n, p^c_n) = H(\{\lambda^s, \lambda^c\}) + \sum_{z:p(z)>0} p(z)\\
\{-H(\{\lambda^{s|\pi^*_z}, \lambda^{c|\pi^*_z}\}) + D(p^{s|\pi^*_z}, p^{c|\pi^*_z})\},\\
= 0.7219 - 0.1176 = 0.6043
\end{aligned}$$

Remark 1: Note that the ability to quantify the secrecy loss may provide a good indication of how secure the system is; for instance for the above example, 0.6043 may or may not be an acceptable level of secrecy loss depending on the application. For example cache memory side channel attacks were explored in [20]. Our computation shows that no amount of secret could be revealed through the side-channel if the cache line is periodically evicted by the processor.

VI. Conclusion

In this paper we presented a quantification of the level of information-flow secrecy loss in partially-observed stochastic systems modeled as partially-observed labeled Markov chains, where information about the system secrets may be revealed through the side-channel of observable inputs/outputs. The statistical difference between the influence of secrets versus covers on the observables, in the form of the Jensen Shannon Divergence measure, is employed to quantify the loss of secrecy. We proposed the computation of the "limiting" JSD as a measure of worst case secrecy loss, resulting from their longer and longer observations. We also presented a state-based approach for computing the fixed-point or limiting JSD. The computation of JSD is challenging since a finite-state Markov chain under partial observations is potentially infinite-state (with the state-space being the conditional state distributions following the observations), and while a finite-state observer model is possible, the model is no longer a Markov chain (since it does not possess scalar transition probabilities).

References

[1] D. Kundur and K. Ahsan, "Practical internet steganography: Data hiding in ip," in *Proc. Texas Workshop on Security of Information Systems*, College Station, Texas, Apr. 2003.

[2] C. T. Christian S. Collberg, "Watermarking, tamper-proofing, and obfuscation-tools for software protection," *IEEE trans. Software Engineering*, vol. 28, no. 8, pp. 735–746, Aug. 2002.

[3] J. Ren and J. Wu, "Survey on anonymous communications in computer networks," *Computer Communications*, vol. 33, pp. 420–431, 2010.

[4] S. Garg, C. Gentry, S. Halevi, M. Raykova, A. Sahai, and B. Waters, "Candidate indistinguishability obfuscation and functional encryption for all circuits," in *IEEE 54th Annual Symposium on Foundations of Computer Science (FOCS' 2013)*, Berkeley, CA, Oct. 2013, pp. 40–49.

[5] 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.

[6] M. Backes, B. Köpf, and A. Rybalchenko, "Automatic discovery and quantification of information leaks," in *Proc. 30th IEEE Symp. Security and Privacy*, Washington, DC, May 2009, pp. 141–153.

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

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] M. A. Ré and R. K. Azad, "Generalization of entropy based divergence measures for symbolic sequence analysis," *PLoS ONE*, vol. 9, no. (4):e93532, Apr. 2014.

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

[16] 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.

[17] 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.

[18] 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.

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

[20] T. Zhang and ruby B. Lee, "Secure cache modeling for measuring side-channel leakage," *Technical Report, Princeton University*, 2014.