Information Flow in Interactive Systems

Mário S. Alvim[1], Miguel E. Andrés[2], and Catuscia Palamidessi[1].

[1]INRIA and LIX, École Polytechnique Palaiseau, France.
[2]Institute for Computing and Information Sciences, The Netherlands.

Abstract. We consider the problem of defining the information leakage in interactive systems where secrets and observables can alternate during the computation. We show that the information-theoretic approach which interprets such systems as (simple) noisy channels is not valid anymore. However, the principle can be recovered if we consider more complicated types of channels, that in Information Theory are known as channels with memory and feedback. We show that there is a complete correspondence between interactive systems and such kind of channels. Furthermore, we show that the capacity of the channels associated to such systems is a continuous function of the Kantorovich metric.

1 Introduction

Information leakage refers to the problem that the observable parts of the behavior of a system may reveal information that we would like to keep secret. In recent years, there has been a growing interest in the quantitative aspects of this problem, partly because it is convenient to represent the partial knowledge of the secrets as a probability distribution, and partly because the mechanisms to protect the information may use randomization to obfuscate the relation between the secrets and the observables.

Among the quantitative approaches, some of the most popular ones are based on Information Theory [5, 12, 4, 16]. The system is interpreted as an information-theoretic *channel*, where the secrets are the input and the observables are the output. The channel matrix is constituted by the conditional probabilities $p(b \mid a)$, defined as the measure of the executions that give observable b within those which contain the secret a. The leakage is represented by the *mutual information*, and the worst-case leakage by the *capacity* of the channel.

In the above works, the secret value is assumed to be chosen at the beginning of the computation. In this paper, we are interested in *Interactive systems*, i.e. systems in which secrets and observables can alternate during the computation, and influence each other. Examples of interactive protocols include *auction protocols* like [21, 18, 17]. Some of these have become very popular thanks to their integration in Internet-based electronic commerce platforms [9, 10, 14]. As for interactive programs, examples include web servers, GUI applications, and command-line programs [3].

We investigate the applicability of the information-theoretic approach to interactive systems. In [8] it was proposed to define the matrix elements $p(b \mid a)$ as the measure of the traces with (secret, observable)-projection (a, b), divided by the measure of the trace with secret projection a. This follows the definition of conditional probability in terms of joint and marginal probability. However, it does not define an information-theoretic channel. In fact, by definition a channel should be invariant with respect to the input distribution, and such construction is not, as shown by the following example.

Example 1. Figure 1 represents a web-based interaction between one seller and two possible buyers, *rich* and *poor*. The seller offers two different products, *cheap* and *expensive*, with given probabilities. Once the product is offered, each buyer may try to buy the product, with a certain probability. For simplicity we assume that the buyers offers are exclusive. We assume that the offers are observables, in the sense that they are made public in the website, while the identity of the buyer that actually buys the product should be secret to an external observer. The symbols r, s, t, $\overline{r}$, $\overline{s}$, $\overline{t}$ represent the probabilities, with the convention that $\overline{r} = 1 - r$.

Following [8] we can compute the conditional probabilities as $p(b|a) = \frac{p(a,b)}{p(a)}$, thus obtaining the matrix on Table 1.

However, the matrix is not invariant with respect to the input distribution. For instance, if we fix $r = \overline{r} = 0.5$ and consider two different input distributions, obtained by varying the values of (s, t), *we get two different matrices of conditional probabilities*, which are represented in Table 2. Hence when the secrets occur *after* the observables we cannot consider the conditional probabilities as representing a (classical) channel, and we cannot apply the standard information-theoretic concepts. In particular, we cannot adopt the (classical) capacity to represent the worst-case leakage, since the capacity is defined using a fixed channel matrix over all possible input distributions.

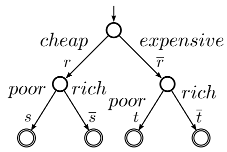


Fig. 1. Inter. System

The first contribution of this paper is to consider an extension of the theory of channels which makes the information-theoretic approach applicable also the case of interactive systems. It turns out that a richer notion of channels, known in Information Theory as *channels with memory and feedback*, serves our purposes. The dependence of inputs on previous outputs corresponds to feedback, and the dependence of outputs on previous inputs and outputs corresponds to memory.

	$cheap$	$expensive$
$poor$	$\frac{rs}{rs+\overline{r}t}$	$\frac{\overline{r}t}{rs+\overline{r}t}$
$rich$	$\frac{r\overline{s}}{r\overline{s}+\overline{r}\overline{t}}$	$\frac{\overline{r}\overline{t}}{r\overline{s}+\overline{r}\overline{t}}$

Table 1. Cond. probabilities of Example 1

	$cheap$	$expensive$	Input dist.
$poor$	$\frac{2}{5}$	$\frac{3}{5}$	$p(poor) = \frac{1}{2}$
$rich$	$\frac{3}{5}$	$\frac{2}{5}$	$p(rich) = \frac{1}{2}$

(a) $r = \frac{1}{2}, s = \frac{2}{5}, t = \frac{3}{5}$

	$cheap$	$expensive$	Input dist.
$poor$	$\frac{1}{4}$	$\frac{3}{4}$	$p(poor) = \frac{1}{5}$
$rich$	$\frac{9}{16}$	$\frac{7}{16}$	$p(rich) = \frac{4}{5}$

(b) $r = \frac{1}{2}, s = \frac{1}{10}, t = \frac{3}{10}$

Table 2. Two different channel matrices induced by two different input distributions

A second contribution of our work is the proof that the channel capacity is a continuous function of the Kantorovich metric on interactive systems. This was pointed out also in [8], however their construction does not work in our case due to the fact that (as far as we understand) it assumes that the probability of a secret action, in any point of the computation, is not 0. This assumption is not guaranteed in our case and therefore we had to proceed differently.

A more complete version of this paper (with proofs) is on line [1].

2 Preliminaries

2.1 Concepts from Information Theory

For more detailed information on this part we refer to [6]. Let A, B denote two random variables with corresponding probability distributions $p_A(\cdot), p_B(\cdot)$, respectively. We shall omit the subscripts when they are clear from the context. Let $\mathcal{A} = \{a_1, \ldots, a_n\}, \mathcal{B} = \{b_1, \ldots, b_m\}$ denote, respectively, the sets of possible values for A and for B.

The *entropy* of A is defined as $H(A) = -\sum_{\mathcal{A}} p(a_i) \log p(a_i)$ and it measures the uncertainty of A. It takes its minimum value $H(A) = 0$ when $p_A(\cdot)$ is a delta of Dirac. The maximum value $H(A) = \log |\mathcal{A}|$ is obtained when $p_A(\cdot)$ is the uniform distribution. Usually the base of the logarithm is set to be 2 and the entropy is measured in *bits*. The *conditional entropy* of A given B is $H(A|B) = -\sum_{\mathcal{B}} p(b_i) \sum_{\mathcal{A}} p(a_j|b_i) \log p(a_j|b_i)$, and it measures the uncertainty of A when B is known. We can prove that $0 \leq H(A|B) \leq H(A)$. The minimum value, 0, is obtained when A is completely determined by B. The maximum value $H(A)$ is obtained when A and B are independent. The *mutual information* between A and B is defined as $I(A;B) = H(A) - H(A|B)$, and it measures the amount of information about A that we gain by observing B. It can be shown that $I(A;B) = I(B;A)$ and $0 \leq I(A;B) \leq H(A)$.

The entropy and mutual information respect the *chain laws*. Namely, given a sequence of random variables $A_1, A_2, \ldots, A_k$ and B, we have:

$$H(A_1, A_2, \ldots, A_k) = \sum_{i=1}^{k} H(A_i|A_1, \ldots, A_{i-1}) \quad (1)$$

$$I(A_1, A_2, \ldots, A_k; B) = \sum_{i=1}^{k} I(A_i; B|A_1, \ldots, A_{i-1}) \quad (2)$$

A (*discrete memoryless*) *channel* is a tuple $(\mathcal{A}, \mathcal{B}, p(\cdot|\cdot))$, where $\mathcal{A}, \mathcal{B}$ are the sets of input and output symbols, respectively, and $p(b_j|a_i)$ is the probability of observing the output symbol b_j when the input symbol is a_i. An input distribution $p(a_i)$ over $\mathcal{A}$ determines, together with the channel, the joint distribution $p(a_i, b_j) = p(a_i|b_j) \cdot p(a_i)$ and consequently $I(A;B)$. The maximum $I(A;B)$ over all possible input distributions is the channel's *capacity*. Shannon's famous result states that the capacity coincides with the maximum rate by which information can be transmitted using the channel.

In this paper we consider input and output *sequences* instead of just symbols.

Convention 1. *Let $\mathcal{A} = \{a_1, \ldots, a_n\}$ be a finite set of n different symbols (*alphabet*). When we have a sequence of symbols (ordered in time), we use a Greek letter α_t to denote the symbol at time t. The notation α^t stands for the sequence $\alpha_1 \alpha_2 \ldots \alpha_t$. For instance, in the sequence $a_3 a_7 a_5$, we have $\alpha_2 = a_7$ and $\alpha^2 = a_3 a_7$.*

Convention 2. *Let X be a random variable. X^t denotes the sequence of t consecutive occurrences $X_1, \ldots, X_t$ of the random variable X.*

When the channel is used repeatedly, the discrete memoryless channel described above represents the case in which the behavior of the channel at the present time does not depend upon the past history of inputs and outputs. If this assumption does not hold, then we have a channel *with memory*. Furthermore, if the outputs from the channel can be fed back to the encoder, thus influencing the generation of the next input symbol, then the channel is said to be *with feedback*; otherwise it is *without feedback*.

Equation 3 makes explicit the probabilistic behavior of channels regarding those classifications. Suppose a general channel from $\mathcal{A}$ to $\mathcal{B}$ with the associated random variables A for input and B for output. Using the notation introduced in Convention 1, the channel behavior after T uses can be fully described by the joint probability $p(\alpha^T, \beta^T)$.

Using probability laws we derive:

$$p(\alpha^T, \beta^T) = \prod_{t=1}^{T} p(\alpha_t|\alpha^{t-1}, \beta^{t-1}) p(\beta_t|\alpha^t, \beta^{t-1}) \qquad \text{(by the expansion law)} \quad (3)$$

The first term $p(\alpha_t|\alpha^{t-1}, \beta^{t-1})$ indicates that the probability of α_t depends not only on α^{t-1}, but also on β^{t-1} (*feedback*). The second term $p(\beta_t|\alpha^t, \beta^{t-1})$ indicates that the probability of each β_t depends on previous history of inputs α^t and outputs β^{t-1} (*memory*).

If the channel is without feedback, then we have that $p(\alpha_t|\alpha^{t-1}, \beta^{t-1}) = p(\alpha_t|\alpha^{t-1})$, and if the channel is without memory, then we have also $p(\beta_t|\alpha^t, \beta^{t-1}) = p(\beta_t|\alpha_t)$. From these we derive $p(\beta^T|\alpha^T) = \prod_{t=1}^{T} p(\beta_t|\alpha_t)$, which is the classic equation for discrete memoryless channels without feedback.

Let $(\mathcal{V}, \mathcal{K})$ be a Borel space and let $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ and $(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$ be Polish spaces equipped with their Borel σ-algebras. Let $\rho(\mathrm{d}x|v)$ be a family of measures on $\mathcal{X}$ given $\mathcal{V}$. Then $\rho(\mathrm{d}x|v)$ is a *stochastic kernel* if and only if and only if $\rho(\cdot|v)$ is a random variable from $\mathcal{V}$ into the power set $\mathcal{P}(\mathcal{X})$.

2.2 Probabilistic automata

A function $\mu\colon \mathcal{S} \to [0,1]$ is a *discrete probability distribution* on a countable set $\mathcal{S}$ if $\sum_{s\in\mathcal{S}} \mu(s) = 1$ and $\mu(s) \geq 0$ for all s. The set of all discrete probability distributions on $\mathcal{S}$ is $\mathcal{D}(\mathcal{S})$.

A *probabilistic automaton* [15] is a quadruple $M = (\mathcal{S}, \mathcal{L}, \hat{s}, \vartheta)$ where $\mathcal{S}$ is a countable set of *states*, $\mathcal{L}$ a finite set of *labels* or *actions*, $\hat{s}$ the *initial* state, and ϑ a *transition function* $\vartheta : \mathcal{S} \to \wp_f(\mathcal{D}(\mathcal{L} \times \mathcal{S}))$. Here $\wp_f(X)$ is the set of all finite subsets of X. If $\vartheta(s) = \emptyset$ then s is a *terminal* state. We write $s{\to}\mu$ for $\mu \in \vartheta(s),\ s \in \mathcal{S}$. Moreover, we write $s\xrightarrow{\ell}r$ for $s, r \in \mathcal{S}$ whenever $s{\to}\mu$ and $\mu(\ell, r) > 0$. A *fully probabilistic automaton* is a probabilistic automaton satisfying $|\vartheta(s)| \leq 1$ for all states. When $\vartheta(s) \neq \emptyset$ we overload the notation and denote $\vartheta(s)$ the distribution outgoing from s.

A *path* in a probabilistic automaton is a sequence $\sigma = s_0 \xrightarrow{\ell_1} s_1 \xrightarrow{\ell_2} \cdots$ where $s_i \in \mathcal{S}$, $\ell_i \in \mathcal{L}$ and $s_i \xrightarrow{\ell_{i+1}} s_{i+1}$. A path can be *finite* in which case it ends with a state.

A path is *complete* if it is either infinite or finite ending in a terminal state. Given a finite path σ, $\mathrm{last}(\sigma)$ denotes its last state. Let $\mathrm{Paths}_s(M)$ denote the set of all paths, $\mathrm{Paths}^\star{}_s(M)$ the set of all finite paths, and $\mathrm{CPaths}_s(M)$ the set of all complete paths of an automaton M, starting from the state s. We will omit s if $s = \hat{s}$. Paths are ordered by the prefix relation, which we denote by $\leq$. The *trace* of a path is the sequence of actions in $\mathcal{L}^* \cup \mathcal{L}^\infty$ obtained by removing the states, hence for the above σ we have $\mathit{trace}(\sigma) = l_1 l_2 \ldots$. If $\mathcal{L}' \subseteq \mathcal{L}$, then $\mathit{trace}_{\mathcal{L}'}(\sigma)$ is the projection of $\mathit{trace}(\sigma)$ on the elements of $\mathcal{L}'$.

Let $M = (\mathcal{S}, \mathcal{L}, \hat{s}, \vartheta)$ be a (fully) probabilistic automaton, $s \in \mathcal{S}$ a state, and let $\sigma \in \mathrm{Paths}^\star_s(M)$ be a finite path starting in s. The *cone* generated by σ is the set of complete paths $\langle\sigma\rangle = \{\sigma' \in \mathrm{CPaths}_s(M) \mid \sigma \leq \sigma'\}$. Given a fully probabilistic automaton $M = (\mathcal{S}, \mathcal{L}, \hat{s}, \vartheta)$ and a state s, we can calculate the *probability value*, denoted by $\mathbf{P}_s(\sigma)$, of any finite path σ starting in s as follows: $\mathbf{P}_s(s) = 1$ and $\mathbf{P}_s(\sigma \xrightarrow{\ell} s') = \mathbf{P}_s(\sigma)\, \mu(\ell, s')$, where $\mathrm{last}(\sigma) \rightarrow \mu$.

Let $\Omega_s \triangleq \mathrm{CPaths}_s(M)$ be the sample space, and let $\mathcal{F}_s$ be the smallest σ-algebra generated by the cones. Then $\mathbf{P}$ induces a unique *probability measure* on $\mathcal{F}_s$ (which we will also denote by $\mathbf{P}_s$) such that $\mathbf{P}_s(\langle\sigma\rangle) = \mathbf{P}_s(\sigma)$ for every finite path σ starting in s. For $s = \hat{s}$ we write $\mathbf{P}$ instead of $\mathbf{P}_{\hat{s}}$.

Given a probability space $(\Omega, \mathcal{F}, P)$ and two events $A, B \in \mathcal{F}$ with $P(B) > 0$, the *conditional probability* of A given B, $P(A \mid B)$, is defined as $P(A \cap B)/P(B)$.

3 Discrete channels with memory and feedback

We adopt the model proposed in [19] for discrete channels with memory and feedback. Such model, represented in Figure 2, can be decomposed in sequential components as follows. At time t the internal channel's behavior is represented by the conditional probabilities $p(\beta_t|\alpha^t, \beta^{t-1})$. The internal channel takes the input α_t and, according to the history of inputs and outputs up to the moment α^t, β^{t-1}, produces an output symbol β_t. The output is then fed back to the encoder with delay one. On the other side, at time t the encoder takes the message and the past output symbols β^{t-1}, and produces a channel input symbol α_t. At final time T the decoder takes all the channel outputs β^T and produces the decoded message $\hat{W}$. The order is the following:

$$\text{Message } W, \quad \alpha_1, \beta_1, \quad \alpha_2, \beta_2, \quad \ldots, \quad \alpha_T, \beta_T, \quad \text{Decoded Message } \hat{W}$$

Let us describe such channel in more detail. Let $\mathcal{A}$ and $\mathcal{B}$ be two finite sets. Let $\{A_t\}_{t=1}^T$ (channel's input) and $\{B_t\}_{t=1}^T$ (channel's output) be families of random variables in $\mathcal{A}$ and $\mathcal{B}$ respectively. Moreover, let $\mathcal{A}^T$ and $\mathcal{B}^T$ represent their T-fold product spaces. A *channel* is a family of stochastic kernels $\{p(\beta_t|\alpha^t, \beta^{t-1})\}_{t=1}^T$.

Let $\mathcal{F}_t$ be the set of all measurable maps $\varphi_t : \mathcal{B}^{t-1} \rightarrow \mathcal{A}$ endowed with a probability distribution, and let F_t be the corresponding random variable. Let $\mathcal{F}^T$, F^T denote the Cartesian product on the domain and the random variable, respectively. A *channel code function* is an element $\varphi^T = (\varphi_1, \ldots, \varphi_T) \in \mathcal{F}^T$.

Note that, by probability laws, $p(\varphi^T) = \prod_{t=1}^T p(\varphi_t|\varphi^{t-1})$. Hence the distribution on $\mathcal{F}^T$ is uniquely determined by a sequence $\{p(\varphi_t|\varphi^{t-1})\}_{t=1}^T$. We will use the notation $\varphi^t(\beta^{t-1})$ to represent the $\mathcal{A}$-valued t-tuple $(\varphi_1, \varphi_2(\beta^1), \ldots, \varphi_t(\beta^{t-1}))$.

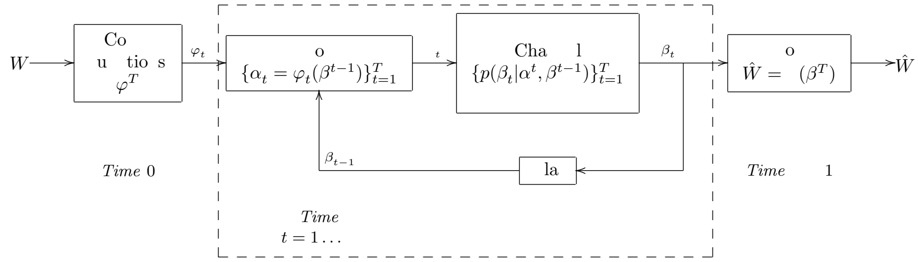


Fig. 2. Model for discrete channel with memory and feedback

In Information Theory this kind of channels are used to encode and transmit messages. If $\mathcal{W}$ is a message set of cardinality M with typical element w, endowed with a probability distribution, a *channel code* is a set of M channel code functions $\varphi^T[w]$, interpreted as follows: for message w, if at time t the channel feedback is β^{t-1}, then the channel encoder outputs $\varphi_t[w](\beta^{t-1})$. A *channel decoder* is a map from $\mathcal{B}^T$ to $\mathcal{W}$ which attempts to reconstruct the input message after observing all the output history β^T from the channel.

3.1 Directed information and capacity of channels with feedback

In classical Information Theory, the channel capacity, which is related to the channel's transmission rate by Shannon's fundamental result, can be obtained as the supremum of the mutual information over all possible input's distributions. In presence of feedback, however, this correspondence does not hold anymore. More specifically, mutual information does not represent any longer the information flow from α^T to β^T. Intuitively, this is due to the fact that mutual information expresses correlation, and therefore it is increased by feedback. But the feedback, i.e the way the output influences the next input, is part of the a priori knowledge, and therefore should not be counted when we measure the output's contribution to the reduction of the uncertainty about the input. If we want to maintain the correspondence with the transmission rate and with information flow, we need to replace mutual information with *directed information* [13].

Definition 1. *In a channel with feedback, the directed information from input A^T to output B^T is defined as $I(A^T \rightarrow B^T) = \sum_{t=1}^T I(\alpha^t; \beta_t|\beta^{t-1})$. In the other direction, the directed information from B^T to A^T is defined as: $I(B^T \rightarrow A^T) = \sum_{t=1}^T I(\alpha_t; \beta^{t-1}|\alpha^{t-1})$.*

Note that the directed information defined above are not symmetric: the flow from A^T to B^T takes into account the correlation between α^t and β_t, while the flow from B^T to A^T is based on the correlation between β^{t-1} and α_t . Intuitively, this is because α^t influences β_t, but, in the other direction, it is β^{t-1} that influences α_t.

It can be proved [19] that $I(A^T; B^T) = I(A^T \rightarrow B^T) + I(B^T \rightarrow A^T)$. If a channel does not have feedback, then $I(B^T \rightarrow A^T) = 0$ and $I(A^T; B^T) = I(A^T \rightarrow B^T)$.

In a channel with feedback the information transmitted is the directed information, and not the mutual information. The following example should help understanding why.

Example 2. Consider the discrete memoryless channel with input alphabet $\mathcal{A} = \{a_1, a_2\}$ and output alphabet $\mathcal{B} = \{b_1, b_2\}$ whose matrix is represented in Table 3.

	b_1	b_2
a_1	0.5	0.5
a_2	0.5	0.5

Table 3. Channel matrix for Example 2

Suppose that the channel is used with feedback, in such a way that, for all t's, $\alpha_{t+1} = a_1$ if $\beta_t = b_1$, and $\alpha_{t+1} = a_2$ if $\beta_t = b_2$. It is easy to show that if $t \geq 2$ then $I(A^t; B^t) \neq 0$. However, there is no leakage from from A^t to B^t, since the rows of the matrix are all equal. We have indeed that $I(A^t \rightarrow B^t) = 0$, and the mutual information $I(A^t; B^t)$ is only due to the feedback information flow $I(B^t \rightarrow A^t)$.

The concept of capacity is generalized for channels with feedback as follows. Let $\mathcal{D}_T = \{\{p(\alpha_t|\alpha^{t-1}, \beta^{t-1})\}_{t=1}^T\}$ be the set of all input distributions. For finite T, the capacity of a channel $\{p(\beta_t|\alpha^t, \beta^{t-1})\}_{t=1}^T$ is:

$$C_T = \sup_{\mathcal{D}_T} \frac{1}{T} I(A^T \rightarrow B^T) \tag{4}$$

4 Interactive systems as channels with memory and feedback

(General) Interactive Information Hiding Systems ([2]), are a variant of probabilistic automata in which we separate actions in secret and observable; "interactive" means that secret and observable actions can interleave and influence each other.

Definition 2. *A general* IIHS *is a quadruple* $\mathcal{I} = (M, \mathcal{A}, \mathcal{B}, \mathcal{L}_\tau)$*, where* M *is a probabilistic automaton* $(\mathcal{S}, \mathcal{L}, \hat{s}, \vartheta)$*,* $\mathcal{L} = \mathcal{A} \cup \mathcal{B} \cup \mathcal{L}_\tau$ *where* $\mathcal{A}$*,* $\mathcal{B}$*, and* $\mathcal{L}_\tau$ *are pairwise disjoint sets of secret, observable, and internal actions respectively, and* $\vartheta(s) \subseteq \mathcal{D}(\mathcal{B} \cup \mathcal{L}_\tau \times \mathcal{S})$ *implies* $|\vartheta(s)| \leq 1$*, for all* s*. The condition on* ϑ *ensures that all observable transitions are fully probabilistic.*

Assumption In this paper we assume that general IIHSs are *normalized*, i.e. once unfolded, all the transitions between two consecutive levels have either secret labels only, or observable labels only. Moreover, the occurrences of secret and observable labels alternate between levels. We will call *secret states* the states from which only secrets-labeled transitions are possible, and *observable states* the others. Finally, we assume that for every s and ℓ there exists a unique r such that $s \xrightarrow{\ell} r$. Under this assumption we have that the traces of a computation determine the final state, as expressed by the next proposition. In the following $trace_{\mathcal{A}}$ and $trace_{\mathcal{B}}$ indicate the projection of the traces on secret and observable actions, respectively. Given a general IIHS, it is always possible to find an equivalent one that satisfies this assumptions. The interested reader can find in [1] the formal definition of the transformation.

Proposition 1. *Let* $\mathcal{I} = (M, \mathcal{A}, \mathcal{B}, \mathcal{L}_\tau)$ *be a general* IIHS*. Consider two paths* σ *and* σ'*. Then,* $trace_{\mathcal{A}}(\sigma) = trace_{\mathcal{A}}(\sigma')$ *and* $trace_{\mathcal{B}}(\sigma) = trace_{\mathcal{B}}(\sigma')$ *implies* $\sigma = \sigma'$*.*

In the following, we will consider two particular cases: the *fully probabilistic* IIHSs, where there is no nondeterminism, and the *secret -nondeterministic* IIHSs, where each secret choice is fully nondeterministic. The latter will be called simply IIHSs.

Definition 3. *Let $\mathcal{I} = ((\mathcal{S}, \mathcal{L}, \hat{s}, \vartheta), \mathcal{A}, \mathcal{B}, \mathcal{L}_\tau)$ be a general* IIHS. *Then $\mathcal{I}$ is:*

- *fully probabilistic if $\vartheta(s) \subseteq \mathcal{D}(\mathcal{A} \times \mathcal{S})$ implies $|\vartheta(s)| \leq 1$ for each $s \in \mathcal{S}$.*
- *secret-nondeterministic if $\vartheta(s) \subseteq \mathcal{D}(\mathcal{A} \times \mathcal{S})$ implies that for each $s \in \mathcal{S}$ there exist s_i' such that $\vartheta(s) = \{\delta(a_i, s_i)\}_{i=1}^n$.*

We show now how to construct a channel with memory and feedback from IIHSs. We will see that an IIHS corresponds precisely to a channel as determined by its stochastic kernel, while a fully probabilistic IIHS determines, additionally, the input distribution. In the following, we consider an IIHS $\mathcal{I} = ((\mathcal{S}, \mathcal{L}, \hat{s}, \vartheta), \mathcal{A}, \mathcal{B}, \mathcal{L}_\tau)$ is in *normalized form*. Given a path σ of length $2t-1$, we denote $trace_{\mathcal{A}}(\sigma)$ by α^t, and $trace_{\mathcal{B}}(\sigma)$ by β^{t-1}.

Definition 4. *For each t, the channel's stochastic kernel corresponding to $\mathcal{I}$ is defined as $p(\beta_t|\alpha^t, \beta^{t-1}) = \vartheta(q)(\beta_t, q')$, where q is the state reached from the root via the path σ whose input-trace is α^t and output trace β^{t-1}.*

Note that q and q' in previous definitions are well defined: by Proposition 1, q is unique, and since the choice of β_t is fully probabilistic, q' is also unique.

If $\mathcal{I}$ is fully probabilistic, then it determines also the input distribution and the dependency of α_t upon β^{t-1} (feedback) and α^{t-1}.

Definition 5. *If $\mathcal{I}$ is fully probabilistic, the associated channel has a conditional input distribution for each t defined as $p(\alpha_t|\alpha^{t-1}, \beta^{t-1}) = \vartheta(q)(\alpha_t, q')$, where q is the state reached from the root via the path σ whose input-trace is α^{t-1} and output trace is β^{t-1}.*

4.1 Lifting the channel inputs to reaction functions

Definitions 4 and 5 define the joint probabilities $p(\alpha^t, \beta^t)$ for a fully probabilistic IIHS. We still need to show in what sense these define a information-theoretic channel.

The $\{p(\beta_t|\alpha^t, \beta^{t-1})\}_{t=1}^T$ determined by the IIHS correspond to a channel's stochastic kernel. The problem resides in the conditional probability of $\{p(\alpha_t|\alpha^{t-1}, \beta^{t-1})\}_{t=1}^T$. In an information-theoretic channel, the value of α_t is determined in the encoder by a deterministic function $\varphi_t(\beta^{t-1})$. However, inside the encoder there is no possibility for a probabilistic description of α_t. Furthermore, in our setting the concept of encoder makes no sense as there is no information to encode. A solution to this problem is to externalize the probabilistic behavior of α_t: the code functions become simple *reaction functions* φ_t that depend only on β^{t-1} (the message w does not play a role any more), and these reaction functions are endowed with a probability distribution that generates the probabilistic behavior of the values of α_t.

Definition 6. A reactor *is a distribution on reaction functions, i.e., a stochastic kernel $\{p(\varphi_t|\varphi^{t-1})\}_{t=1}^T$. A reactor R is* consistent with a fully probabilistic IIHS $\mathcal{I}$ *if it induces the compatible distribution $Q(\varphi^T, \alpha^T, \beta^T)$ such that, for every $1 \leq t \leq T$, $Q(\alpha_t|\alpha^{t-1}, \beta^{t-1}) = p(\alpha_t|\alpha^{t-1}, \beta^{t-1})$, where the latter is the probability distribution induced by $\mathcal{I}$.*

The main result of this section states that for any fully probabilistic IIHS there is a reactor that generates the probabilistic behavior of the IIHS.

Theorem 3. *Given a fully probabilistic* IIHS $\mathfrak{I}$ *, we can construct a channel with memory and feedback, and probability distribution* $Q(\varphi^T, \alpha^T, \beta^T)$*, which corresponds to* $\mathfrak{I}$ *in the sense that, for every* t*,* α^t *and* β^t*, with* $1 \leq t \leq T$*,* $Q(\alpha^t, \beta^t) \stackrel{\text{def}}{=} \sum_{\varphi^T} Q(\varphi^T, \alpha^t, \beta^t) = p(\alpha^t, \beta^t)$ *holds, where* $p(\alpha^t, \beta^t)$ *is the joint probability of input and output traces induced by* $\mathfrak{I}$*.*

Corollary 1. *Let a* $\mathfrak{I}$ *be a fully probabilistic* IIHS*. Let* $\{p(\beta_t|\alpha^t, \beta^{t-1})\}_{t=1}^T$ *be a sequence of stochastic kernels and* $\{p(\alpha_t|\alpha^{t-1}, \beta^{t-1})\}_{t=1}^T$ *a sequence of input distributions defined by* $\mathfrak{I}$ *according to Definitions 4 and 5. Then the reactor* $R = \{p(\varphi_t|\varphi^{t-1})\}_{t=1}^T$ *compatible with respect to the* $\mathfrak{I}$ *is given by:*

$$p(\varphi_1) = p(\alpha_1|\alpha^0, \beta^0) = p(\alpha_1) \tag{5}$$

$$p(\varphi_t|\varphi^{t-1}) = \prod_{\beta^{t-1}} p(\varphi_t(\beta^{t-1})|\varphi^{t-1}(\beta^{t-2}), \beta^{t-1}), \quad 2 \leq t \leq T \tag{6}$$

Figure 3 depicts the model for IIHS. Note that, in relation to Figure 2, there are some simplifications: (1) no message w is needed; (2) the decoder is not used. At the beginning, a reaction function sequence φ^T is chosen and then the channel is used T times. At each usage t, the encoder decides the next input symbol α_t based on the reaction function φ_t and the output fed back β^{t-1}. Then the channel produces an output β_t based on the stochastic kernel $p(\beta_t|\alpha^t, \beta^{t-1})$. The output is then fed back to the encoder with a delay one.

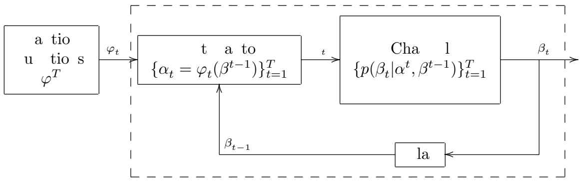


Fig. 3. Channel with memory and feedback model for IIHS

We conclude this section by remarking an intriguing coincidence: The notion of reaction function sequence φ^T, on the IIHSs, corresponds to the notion of deterministic scheduler. In fact, each reaction function φ_t selects the next step, α_t, on the basis of the β^{t-1} and α^{t-1} (generated by φ^{t-1}), and β^{t-1}, α^{t-1} represent the path until that state.

5 Leakage in Interactive Systems

In this section we propose a notion of information flow based on our model. We follow the idea of defining leakage and maximum leakage using the concepts of mutual information and capacity (see for instance [4]), making the necessary adaptations.

Since the directed information $I(A^T \to B^T)$ is a measure of how much information flows from A^T to B^T in a channel with feedback (cfr. Section 3.1), it is natural to consider it as a measure of leakage of information by the protocol.

Definition 7. *The information leakage of an* IIHS *is defined as:* $I(A^T \to B^T) = \sum_{t=1}^{T} H(A_t|A^{t-1}, B^{t-1}) - H(A^T|B^T)$.

Note that $\sum_{t=1}^{T} H(A_t|A^{t-1}, B^{t-1})$ can be seen as the entropy H_R of reactor R.

Compare this definition with the classical Information-theoretic approach to information leakage: when there is no feedback, the leakage is defined as:

$$I(A^T; B^T) = H(A^T) - H(A^T|B^T) \tag{7}$$

The principle behind (7) is that the leakage is equal to the difference between the *a priori uncertainty* $H(A^T)$ and the *a posteriori uncertainty* $H(A^T|B^T)$ (gain in knowledge about the secret by observing the output). Our definition maintains the same principle, with the proviso that the a priori uncertainty is now represented by H_R.

5.1 Maximum leakage as capacity

In the case of secret-nondeterministic IIHS, we have a stochastic kernel but no distribution on the code functions. In this case it seems natural to consider the worst leakage over all possible distributions on code functions. This is exactly the concept of capacity.

Definition 8. *The* maximum leakage *of an* IIHS *is defined as the capacity* C_T *of the associated channel with memory and feedback.*

6 Modeling IIHSs as channels: An example

In this section we show the application of our approach to the *Cocaine Auction Protocol* [17]. Let us imagine a situation where several mob individuals are gathered around a table. An auction is about to be held in which one of them offers his next shipment of cocaine to the highest bidder. The seller describes the merchandise and proposes a starting price. The others then bid increasing amounts until there are no bids for 30 consecutive seconds. At that point the seller declares the auction closed and arranges a secret appointment with the winner to deliver the goods.

The basic protocol is fairly simple and is organized as a succession of rounds of bidding. Round i starts with the seller announcing the bid price b_i for that round. Buyers have t seconds to make an offer (i.e. to say yes, meaning "I'm willing to buy at the current bid price b_i"). As soon as one buyer anonymously says yes, he becomes the winner w_i of that round and a new round begins. If nobody says anything for t seconds, round i is concluded by timeout and the auction is won by the winner w_{i-1} of the previous round, if one exists. If the timeout occurs during round 0, this means that nobody made any offers at the initial price b_0, so there is no sale.

Although our framework allows the forrmalization of this protocol for an arbitrary number of bidders and bidding rounds, for illustration purposes, we will consider the case of two bidders (*Candlemaker* and *Scarface*) and two rounds of bids. Furthermore, we assume that the initial bid is always 1 dollar, so the first bid does not need to be announced by the seller. In each turn the seller can choose how much he wants to increase the actual bid. This is done by adding an increment to the last bid. There are two options of increments, namely inc_1 (1 dollar) and inc_2 (2 dollars). In that way, b_{i+1} is either $b_i + inc_1$ or $b_i + inc_2$. We can describe this protocol as a *normalized* IIHS $\mathcal{I} = (M, \mathcal{A}, \mathcal{B}, \mathcal{L}_\tau)$, where $\mathcal{A} = \{\text{Candlemaker, Scarface}, a^*\}$ is the set of secret actions, $\mathcal{B} = \{inc_1, inc_2, b_*\}$ is the set of observable actions, $\mathcal{L}_\tau = \emptyset$ is the set of hidden actions, and the probabilistic automaton M is represented in Figure 4. For clarity reasons, we omit transitions with probability 0 in the automaton. Note that the special secret action a_* represents the situation where neither *Candlemaker* nor *Scarface* bid. The special observable action b_* is only possible after no one has bidden, and signalizes the end of the auction and, therefore, no bid is allowed anymore.

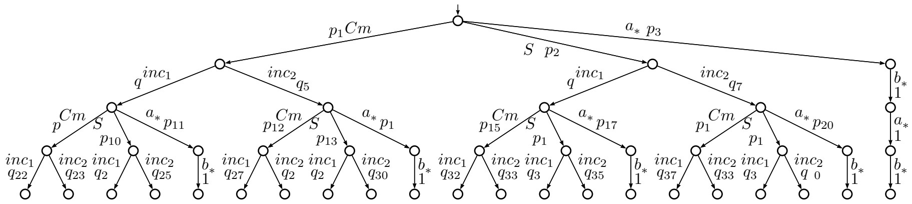


Fig. 4. Cocaine Auction example

Table 4 shows all the stochastic kernels for this example. The formalization of this protocol in terms of IIHSs using our framework makes it possible to prove the claim in[17] suggesting that if the seller knows the identity of the bidders then the (strong) anonymity guaranties are not provided anymore.

7 Topological properties of IIHSs and their Capacity

In this section we show how to extend to IIHSs the notion of pseudometric defined in [8] for Concurrent Labelled Markov Chains, and we prove that the capacity of the corresponding channels is a continuous function on this pseudometric. The metric construction is sound for general IIHSs, but the result on capacity is only valid for secret-nondeterministic IIHSs.

Given a set of states S, a pseudometric (or distance) is a function d that yields a non-negative real number for each pair of states and satisfies the following: $d(s, s) = 0$; $d(s, t) = d(t, s)$, and $d(s, t) \leq d(s, u) + d(u, t)$. We say that a pseudometric d is c-bounded if $\forall s, t : d(s, t) \leq c$, where c is a positive real number. We now define a complete lattice on pseudometrics, and define the distance between IIHSs as the greatest fixpoint of a distance transformation, in line with the coinductive theory of bisimilarity.

$\alpha_1 \to \beta_1$	inc_1	inc_2	b_*
Candlemaker	q_4	q_5	0
Scarface	q_6	q_7	0
a^*	0	0	1

(a) $t=1, p(\beta_1|\alpha^1, \beta^0)$

$\alpha_1, \beta_1, \alpha_2 \to \beta_2$	Cheap	Expensive	b_*
Candlemaker,inc_1,*Candlemaker*	q_{22}	q_{23}	0
Candlemaker,inc_1,*Scarface*	q_{24}	q_{25}	0
Candlemaker,inc_1,a_*	0	0	1
Candlemaker,inc_2,*Candlemaker*	q_{27}	q_{28}	0
Candlemaker,inc_2,*Scarface*	q_{29}	q_{30}	0
Candlemaker,inc_2,a_*	0	0	1
Scarface,inc_1,*Candlemaker*	q_{32}	q_{33}	0
Scarface,inc_1,*Scarface*	q_{34}	q_{35}	0
Scarface,inc_1,a_*	0	0	1
Scarface,inc_2,*Candlemaker*	q_{37}	q_{38}	0
Scarface,inc_2,*Scarface*	q_{39}	q_{40}	0
Scarface,inc_2,a_*	0	0	1
a_*,b_*,a_*	0	0	1
All other lines	0	0	1

(b) $t = 2, p(\beta_2|\alpha^2, \beta^1)$

Table 4. Stochastic kernels for the Cocaine Auction example.

Definition 9. *$\mathcal{M}$ is the class of 1-bounded pseudometrics on states with the ordering* $d \preceq d'$ if $\forall s, s' \in S : d(s, s') \geq d'(s, s')$.

It is easy to see that $(\mathcal{M}, \preceq)$ is a complete lattice. In order to define pseudometrics on IIHSs, we now need to lift the pseudometrics on states to pseudometrics on distributions in $\mathcal{D}(\mathcal{L} \times S)$. Following standard lines [20, 8, 7], we apply the construction based on the Kantorovich metric [11].

Definition 10. *For $d \in \mathcal{M}$, and $\mu, \mu' \in \mathcal{D}(\mathcal{L} \times S)$, we define $d(\mu, \mu')$ (overloading the notation d) as $d(\mu, \mu') = \max \sum_{(\ell_i, s_i) \in \mathcal{L} \times S} (\mu(\ell_i, s_i) - \mu'(\ell_i, s_i))x_i$ where the maximization is on all possible values of the x_i's, subject to the constraints $0 \leq x_i \leq 1$ and $x_i - x_j \leq \hat{d}((\ell_i, s_i), (\ell_j, s_j))$, where $\hat{d}((\ell_i, s_i), (\ell_j, s_j)) = 1$ if $\ell_i \neq \ell_j$, and $\hat{d}((\ell_i, s_i), (\ell_j, s_j)) = d(s_i, s_j)$ otherwise.*

It can be shown that with this definition m is a pseudometric on $\mathcal{D}(\mathcal{L} \times S)$.

Definition 11. *$d \in \mathcal{M}$ is a* bisimulation metric *if, for all $\epsilon \in [0, 1)$, $d(s, s') \leq \epsilon$ implies that if $s \to \mu$, then there exists some μ' such that $s' \to \mu'$ and $d(\mu, \mu') \leq \epsilon$.*

The greatest bisimulation metric is $d_{max} = \bigsqcup \{d \in \mathcal{M} \mid d \text{ is a bisimulation metric}\}$. We now characterize d_{max} as a fixed point of a monotonic function Φ on $\mathcal{M}$. For simplicity, from now on we consider only the distance between states belonging to different IIHSs with disjoint sets of states.

Definition 12. *Given two* IIHSs *with transition relations θ and θ' respectively, and a preudometric d on states, define $\Phi : \mathcal{M} \to \mathcal{M}$ as:*

$$\Phi(d)(s,s') = \begin{cases} \max_i d(s_i, s'_i) & \text{if } \quad \vartheta(s) = \{\delta_{(a_1,s_1)}, \ldots, \delta_{(a_m,s_m)}\} \\ & \text{and } \vartheta'(s') = \{\delta_{(a_1,s'_1)}, \ldots, \delta_{(a_m,s'_m)}\} \\ d(\mu,\mu') & \text{if } \vartheta(s) = \{\mu\} \text{ and } \vartheta'(s') = \{\mu'\} \\ 0 & \text{if } \vartheta(s) = \vartheta'(s') = \emptyset \\ 1 & \text{otherwise} \end{cases}$$

It is easy to see that the definition of Φ is a particular case of the function F defined in [8, 7]. Hence it can be proved, by adapting the proofs of the analogous results in [8, 7], that $F(d)$ is a pseudometric, and that d is a bisimulation metric iff $d \preceq \Phi(d)$. This implies that $d_{max} = \bigsqcup\{d \in \mathcal{M} \mid d \preceq \Phi(d)\}$, and still as a particular case of F in [8, 7], we have that Φ is monotonic on $\mathcal{M}$. By Tarski's fixed point theorem, d_{max} is the greatest fixed point of Φ. Furthermore, in [1] we show that d_{max} is indeed a bisimulation metric, and that it is the greatest bisimulation metric. In addition, the finite branchingness of IIHSs ensures that the closure ordinal of Φ is ω (cf. Lemma 3.10 in the full version of [8]). Therefore one can show that $d_{max} = \bigsqcap\{\Phi^i(\top) \mid i \in \mathbb{N}\}$, where $\top$ is the greatest pseudometric (i.e. $\top(s,s') = 0$ for every s, s'), and $\Phi^0(\top) = \top$.

Given two IIHSs $\mathcal{I}$ and $\mathcal{I}'$, with initial states s and s' respectively, we define the distance between $\mathcal{I}$ and $\mathcal{I}'$ as $d(\mathcal{I}, \mathcal{I}') = d_{max}(s, s')$. Next theorem states the continuity of the capacity w.r.t. the metric on IIHSs. It is crucial that they are secret-nondeterministic (while the definition of the metric holds in general).

Theorem 4. *Consider two normalized* IIHSs *$\mathcal{I}$ and $\mathcal{I}'$, and fix a $T > 0$. For every $\epsilon > 0$ there exists $\nu > 0$ such that if $d(\mathcal{I}, \mathcal{I}') < \nu$ then $|C_T(\mathcal{I}) - C_T(\mathcal{I}')| < \epsilon$.*

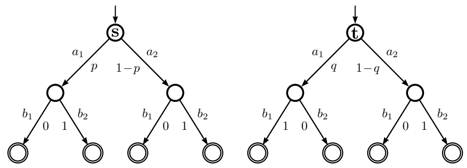
We conclude this section with an example showing that the continuity result for the capacity does not hold if the construction of the channel is done starting from a system in which the secrets are endowed with a probability distribution. This is also the reason why we could not simply adopt the proof technique of the continuity result in [8] and we had to come up with a different reasoning.

Example 3. Consider the two following programs, where a_1, a_2 are secrets, b_1, b_2 are observable, $\|$ is the parallel operator, and $+_p$ is a binary probabilistic choice that assigns probability p to the left branch, and probability $1 - p$ to the right one.

s) $(send(a_1) +_p\ send(a_2)) \parallel receive(x).output(b_2)$
t) $(send(a_1) +_q\ send(a_2)) \parallel receive(x).\mathit{if}\ x = a_1\ \mathit{then}\ output(b_1)\ \mathit{else}\ output(b_2)$.

Table 5 shows the fully probabilistic IIHSs corresponding to these programs, and their associated channels, which in this case (since the secret actions are all at the top-level) are classic channels, i.e. memoryless and without feedback. As usual for classic channels, they do not depend on p and q. It is easy to see that the capacity of the first channel is 0 and the capacity of the second one is 1. Hence their difference is 1, independently from p and q.

Let now $p = 0$ and $q = \epsilon$. It is easy to see that the distance between s and t is ϵ. Therefore (when the automata have probabilities on the secrets), the capacity is not a continuous function of the distance.



s	b_1	b_2
a_1	0	1
a_2	0	1

(a)

t	b_1	b_2
a_1	1	0
a_2	0	1

(b)

Table 5. The IIHSs of Example 3 and their corresponding channels

8 Conclusion and future work

In this paper we have investigated the problem of information leakage in interactive systems, and we have proved that these systems can be modeled as channels with memory and feedback. The situation is summarized in Table 6(a). The comparison with the classical situation of non-interactive systems is represented in (b). Furthermore, we have proved that the channel capacity is a continuous function of the kantorovich metric.

IIHSs **as automata**	IIHSs **as channels**	**Notion of leakage**
Normalized IIHSs with nondeterministic inputs and probabilistic outputs	Sequence of stochastic kernels $\{p(\beta_t \| \alpha^t, \beta^{t-1})\}_{t=1}^T$	Leakage as capacity
Normalized IIHSs with a deterministic scheduler solving the nondeterminism	Sequence of stochastic kernels $\{p(\beta_t \| \alpha^t, \beta^{t-1})\}_{t=1}^T$ + reaction function seq. φ^T	
Fully probabilistic normalized IIHSs	Sequence of stochastic kernels $\{p(\beta_t \| \alpha^t, \beta^{t-1})\}_{t=1}^T$ + reactor $\{p(\varphi_t \| \varphi^{t-1})\}_{t=1}^T$	Leakage as directed information $I(A^T \to B^T)$

(a)

Classical channels	**Channels with memory and feedback**
The protocol is modeled in independent uses of the channel, often a unique use.	The protocol is modeled in several consecutive uses of the channel.
The channel is from $\mathcal{A}^T \to \mathcal{B}^T$, i.e., its input is a single string $\alpha^T = \alpha_1 \dots \alpha_T$ of secret symbols and its output is a single string $\beta^T = \beta_1 \dots \beta_T$ of observable symbols.	The channel is from $\mathcal{F} \to \mathcal{B}$, i.e. its input is a reaction function φ_t and its output is an observable β_t.
The channel is memoryless and in general implicitly it is assumed the absence of feedback.	The channel has memory. Despite the fact that the channel from $\mathcal{F} \to \mathcal{B}$ does not have feedback, the internal stochastic kernels do.
The capacity is calculated using information $I(A^T; B^T)$.	The capacity is calculated using mutual directed information $I(A^T \to B^T)$.

(b)

Table 6.

For future work we would like to provide algorithms to compute the leakage and maximum leakage of interactive systems. These problems result very challenging given the exponential growth of reaction functions (needed to compute the leakage) and the quantification over infinitely many reactors (given by the definition of maximum leakage in terms of capacity). One possible solution is to study the relation between deterministic schedulers and sequence of reaction functions. In particular, we believe that for each sequence of reaction functions and distribution over it there exists a probabilistic scheduler for the automata representation of the secret-nondeterministic IIHS. In this way, the problem of computing the leakage and maximum leakage would reduce to a standard probabilistic model checking problem (where the challenge is to compute probabilities ranging over infinitely many schedulers).

In addition, we plan to investigate measures of leakage for interactive systems other than mutual information and capacity.

References

1. M. S. Alvim, M. E. Andrés, and C. Palamidessi. Information Flow in Interactive Systems, 2010. `http://hal.archives-ouvertes.fr/inria-00479672/en/`.
2. M. E. Andrés, C. Palamidessi, P. van Rossum, and G. Smith. Computing the leakage of information-hiding systems. In *Proc. of TACAS*, volume 6015 of *LNCS*, pages 373–389. Springer, 2010.
3. A. Bohannon, B. C. Pierce, V. Sjöberg, S. Weirich, and S. Zdancewic. Reactive noninterference. In *Proc. of CCS*, pages 79–90. ACM, 2009.
4. K. Chatzikokolakis, C. Palamidessi, and P. Panangaden. Anonymity protocols as noisy channels. *Inf. and Comp.*, 206(2–4):378–401, 2008.
5. D. Clark, S. Hunt, and P. Malacaria. Quantified interference for a while language. In *Proc. of QAPL 2004*, volume 112 of *ENTCS*, pages 149–166. Elsevier, 2005.
6. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. J. Wiley & Sons, Inc., 1991.
7. Y. Deng, T. Chothia, C. Palamidessi, and J. Pang. Metrics for action-labelled quantitative transition systems. In *Proc. of QAPL*, volume 153 of *ENTCS*, pages 79–96. Elsevier, 2006.
8. J. Desharnais, R. Jagadeesan, V. Gupta, and P. Panangaden. The metric analogue of weak bisimulation for probabilistic processes. In *Proc. of LICS*, pages 413–422. IEEE, 2002.
9. Ebay website. `http://www.ebay.com/`.
10. Ebid website. `http://www.ebid.net/`.
11. L. Kantorovich. On the transfer of masses (in Russian). *Doklady Akademii Nauk*, 5(1):1–4, 1942. Translated in *Management Science*, 5(1):1–4, 1958.
12. P. Malacaria. Assessing security threats of looping constructs. In *Proc. of POPL*, pages 225–235. ACM, 2007.
13. J. L. Massey. Causality, feedback and directed information. In *Proc. of the 1990 Intl. Symposium on Information Theory and its Applications*, 1990.
14. Mercadolibre website. `http://www.mercadolibre.com/`.
15. R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, 1995. Tech. Rep. MIT/LCS/TR-676.
16. G. Smith. On the foundations of quantitative information flow. In *Proc. of FOSSACS*, volume 5504 of *LNCS*, pages 288–302. Springer, 2009.
17. F. Stajano and R. J. Anderson. The cocaine auction protocol: On the power of anonymous broadcast. In *Information Hiding*, pages 434–447, 1999.

18. S. Subramanian. Design and verification of a secure electronic auction protocol. In *Proc. of SRDS*, pages 204–210. IEEE, 1998.
19. S. Tatikonda and S. K. Mitter. The capacity of channels with feedback. *IEEE Transactions on Information Theory*, 55(1):323–349, 2009.
20. F. van Breugel and J. Worrell. Towards quantitative verification of probabilistic transition systems. In *Proc. of ICALP*, volume 2076 of *LNCS*, pages 421–432. Springer, 2001.
21. W. Vickrey. Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance*, 16(1):8–37, 1961.