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 [4, 11, 3, 17]. 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 [22, 19, 18]. Some of these have become very popular thanks to their integration in Internet-based electronic commerce platforms [8, 9, 13]. As for interactive programs, examples include web servers, GUI applications, and command-line programs [2].

We investigate the applicability of the information-theoretic approach to interactive systems. In [7] 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 [7] 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, let us assume $r = \overline{r} = \frac{1}{2}$, $s = \frac{2}{3}$, and $t = \frac{2}{3}p$, where p is a parameter. Therefore we have $p(poor) = rs + \overline{r}t = \frac{1}{3}(1 + p)$ or, equivalently, $p = 3 \cdot p(poor) - 1$. Two different input distributions will determine different values of p, and therefore t. Hence also the channel matrices will be different, as the two examples in Table 2 show.

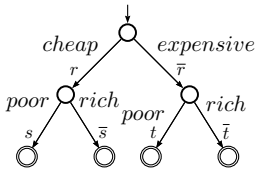


Fig. 1. An interactive syst.

	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. Channel matrix for Example 1

Consequently, 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.

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.

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 [7], 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.

1.1 Plan of the paper

The paper is organized as follows. Section 2 reviews some important concepts fromProbabilistic Automata and Information Theory. It is also presented the concept of Inter-

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

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

	$cheap$	$expensive$	Input distr.
$poor$	$\frac{4}{5}$	$\frac{1}{5}$	$p(poor) = \frac{5}{12}$
$rich$	$\frac{2}{7}$	$\frac{5}{7}$	$p(rich) = \frac{7}{12}$

(b) $r = \frac{1}{2}, s = \frac{2}{3}, p = \frac{1}{4}, t = \frac{1}{6}$

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

active Information Hiding Systems (IIHSs), which will be used all over the paper. In Section 3 we discuss why the classical information theoretical approach needs to be extended and we give an overview on how we do it in our model, along with the main issues involved. Section 4 reviews the model of channels with memory and feedback that are the core the model we propose. The concept of directed information is discussed and also the generalized concept of capacity in the presence of feedback. Section 5 contains our main contribution. We explain how IIHSs can be modeled using channels with memory and feedback. In particular we show that there is always a channel that simulates the probabilistic behavior of any IIHS. In Section 6 we define the quantification of information leakage as the channel's directed information from input to output, in the case where the input distribution on secrets is known, or as the generalized capacity, in the case the distribution on secrets is unknown. In Section 7 we discuss a full example of our model applied to a real protocol. The Cocaine Auction protocol is presented, modeled as a channel with memory and feedback, and then the leakage of information in three different scenarios is calculated. Section 8 discusses the topological properties of IIHSs and their capacity. We show that the capacity of the channels associated to interactive systems is a continuous function of the Kantorovich metric. In Sections 9 and 10 we review and discuss the main results of the paper and illustrate some future work.

A short version of this paper (without proofs, and with less material) appeared in the proceedings of CONCUR 2010.

2 Preliminaries

In this section we briefly review some basic notions that we will need along the paper.

2.1 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 $trace(\sigma) = l_1 l_2 \ldots$. If $\mathcal{L}' \subseteq \mathcal{L}$, then $trace_{\mathcal{L}'}(\sigma)$ is the projection of $trace(\sigma)$ 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 $\mathbf{P}_s(\sigma \xrightarrow{\ell} s') = \mathbf{P}_s(\sigma)\ \mu(\ell, s'),\ \text{where } \mathrm{last}(\sigma) \to \mu$.

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 $P(A \cap B)/P(B)$.

2.2 Concepts from Information Theory

For more detailed information on this part we refer to [5]. 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:

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

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

Let $(\mathcal{V}, \mathcal{K})$ be a set equipped with a σ-algebra of subsets (i.e a Borel space). Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ a Polish space (i.e a separable and completely metrizable topological space), equipped with its Borel σ-algebra. Let $\rho = \{p(\cdot|v)\}_v$ be a family of measures on $\mathcal{X}$ parametrized over $v \in \mathcal{V}$. We say that ρ is a *stochastic kernel* from $\mathcal{V}$ to $\mathcal{X}$ if for every Borel set $B \in \mathcal{B}_\mathcal{X}$, the function $v \mapsto \rho(B|v) \in [0, 1]$ is measurable [1].

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. These conditional probabilities form a stochastic kernel and constitute the *channel matrix*. 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*). We use a Greek letter (α, β, ...) to denote a sequence of symbols (ordered in time). Given a sequence $\alpha = a_{i_1} a_{i_2} \ldots a_{i_m}$, we use α_t to denote the symbol at time t, i.e. a_{i_t}. The notation α^t stands for the sequence $\alpha_{i_1} \alpha_{i_2} \ldots \alpha_{i_t}$. For instance, in the sequence $\alpha = a_3 a_7 a_5$, 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 $p(\alpha^T, \beta^T)$.

[1] For the purpose of this paper, since we only deal with discrete random variables, we only need to assume that for every $v \in \mathcal{V}$, $p(\cdot|v)$ is a probability distribution.

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.

We shall have a deeper discussion about the meaning and implications of Equation (3) later on this paper, when comparing the concepts of mutual information and directed information, in Section 6.

3 Our model vs the classical approach

In this section we illustrate the issues involved in adopting a more general notion of channel, by comparing it with the basic model.

By *classical information theoretical approach*, or simply *classical approach*, we mean the use of discrete memoryless channels, which implicitly assume absence of feedback, to model the problem of information leakage in computational systems. This approach has been used for non-interactive systems, where secrets occur strictly before observables during the computation and therefore do not depend on them. In this paper we extend the classical approach with a richer notion of channels that also consider memory and feedback. Our extension is a generalization of the classical model in the sense that it can represent both interactive and non-interactive systems.

In non-interactive systems, since the secrets always precede the observables, it is possible to group the sequence of secrets (and observables) in a single secret (resp. observable) string. If we consider only one activation of the system, or if each use of the system is independent from the other, then we can model it as a discrete classical channel (memoryless, and without feedback) from a single input string to a single output string.

When we have interactive systems, however, inputs and outputs may interleave and influence each other. Considering some sort of feedback in the channel is a way to capture this richer behavior. Secrets have a causal influence on observables via the channel, and, in the presence of interactivity, observables have a causal influence on secrets via the feedback. This alternating mutual influence between inputs and outputs can be modeled by repeated uses of the channels. However, each time the channel is used it represents a different state of the computation, and the conditional probabilities of observables on secrets can depend on this state. The addition of memory to the model allows expressing the dependency of the channel matrix on such state (which, as we will see, can also be represented by the history of inputs and outputs).

One important feature of the classical approach is that the secret choice is seen as external to the system, i.e. determined by the environment. This implies that the probability distribution on the secrets (input distribution) constitutes the a priori knowledge and does not count as leakage. In order to encompass the classical approach, in our extended model we should preserve this principle, and the most natural way is to consider the secret choices, at every stage of the computation, as external. Their probability distributions, which is now in general a conditional probability distribution (depending on the history of secrets and observables) should be considered as part of the external knowledge, and should not be counted as leakage. This is an important point in our framework, and we wish to draw the attention of the reader on it:

Principle 3. *The probability distributions on the secret choices is part of the external knowledge and it is not considered leakage.*

In Section 9 we will discuss the case in which we remove this assumption, i.e. if we consider leakage also the probabilistic knowledge induced by the distributions on the secret choices.

In summary, the main changes implied by the addition of memory and feedback are:

1. interactive systems are captured by the new model, as well as non-interactive ones as a particular case;
2. in contrast with the usual single use (or independent uses) in the classical approach, the systems behavior is now represented by repeated and dependent uses of the channel;
3. there is a causal relation not only from input to output (via the channel), but also from output to input (via feedback).

Item 3 has a rather strong consequence: In non-interactive systems, only inputs have a causal influence on outputs and mutual information is a good measure of the information flow from secrets to observables. However, in the presence of feedback, outputs also have a causal influence on inputs and although this flow does not correspond to any leakage of secret information (according to Principle 3), it increases mutual information. By definition, indeed, mutual information does not represent causality, but, rather, correlation. In this richer model, mutual information is not a good measure of information leakage anymore. We will come back on this point in Section 4.2, and we will show how to generalize the concept of leakage.

4 Discrete channels with memory and feedback

We adopt the model proposed in [20] 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 time step t, 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:

$$\text{Message } W, \quad \alpha_1, \beta_1, \quad \alpha_2, \beta_2, \quad \ldots, \quad \alpha_T, \beta_T, \quad \text{Decoded Message } \hat{W} \tag{4}$$

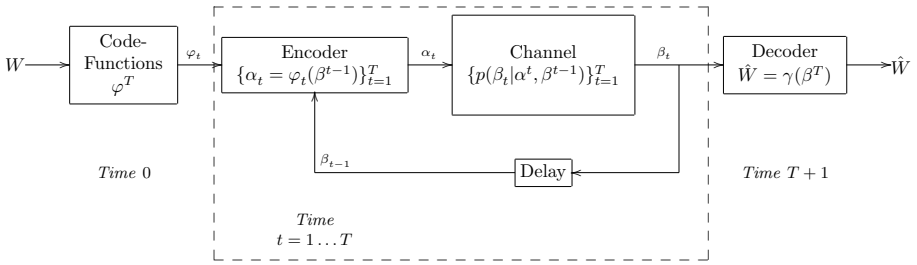


Fig. 2. Model for discrete channel with memory and feedback

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 $\varphi_t : \mathcal{B}^{t-1} \to \mathcal{A}$ 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 $(\varphi_1, \varphi_2(\beta^1), \ldots, \varphi_t(\beta^{t-1}))$.

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.

4.1 The power of memory and feedback

The original purpose of *communication channels* models is to represent data transmission from a source to a receiver. Shannon's famous result states that the maximum information transmission rate with an arbitrarily small probability of error corresponds exactly to the channel capacity. If the number of times the channel is used is large enough, there is an encoding that achieves the optimal transmission rate, i.e the channel capacity. However, Shannon did not explain how to determine such an encoding and,

as a matter of fact, a general way to generate an optimal encoding scheme has not been found yet. But the use of feedback can make the encoding easier, as we shall see now.

Consider a discrete memoryless binary channel $\{\mathcal{A}, \mathcal{B}, p(.|.)\}$ with $\mathcal{A} = \{0, 1\}$ and $\mathcal{B}\{0, 1\}$ and the channel matrix of Table 3. This channel is used to transmit the bits 0 and 1, where the probability of error (i.e transmitting a 1 when the desired bit is 0 or vice-versa) is 0.2. Shannon's theorem guarantees that the maximum information transmission rate in this channel is (2 to the power of) the channel capacity, i.e 0.8 bits per use of the channel.

	0	1
0	0.8	0.2
1	0.2	0.8

Table 3. Channel matrix for binary channel without feedback

The encoding that achieves the capacity can be obtained easily if we add feedback to the channel. Imagine that every bit received on the right hand end of the channel, is feedback noiselessly to the source with delay 1. Define the encoding as follows: for each bit transmitted, the encoder checks via feedback if the received bit was the correct one. If not, the encoder retransmits the bit again and restart the process. If yes, the transmission is considered complete [2].

It is easy to see that with this encoding the transmission rate is 0.8 bit per usage of the channel, since in 80% the cases the bit is transmitted properly, and in 20% a retransmission is needed. Note that the capacity, in this example, was not increased by the use of feedback (it is 0.8 bits with or without feedback). It is a special case of the well known result that *feedback does not increase the capacity of discrete memoryless channels* [5].

An example of channel with memory and feedback. Let us go a bit further with the binary channel example and show how memory and feedback can be used.

Again, the set of possible messages is $\mathcal{W} = \{0, 1\}$. The message W to be transmitted is going to be encoded via code functions into a suitable representation to the stochastic kernel within the channel. The input alphabet is $\mathcal{A} = \{0, 1, X\}$ and the output alphabet is $\mathcal{B} = \{0, 1, X\}$, where X is a special symbol used to mark the end of a successful transmission. We assume that at most T uses of the channel are allowed. We use t, with $1 \leq t \leq T$, to represent the t-th time step.

We assume a simple form of memory, in the form of a dependency of the stochastic kernel on some sort of noise that can vary with time, $\mu(t)$. For the symbol X the transmission is noiseless. More precisely, we assume a stochastic kernel defined as follows:

$$p(\beta_t = 0|\alpha^t = \alpha^{t-1}0, \beta^{t-1}) = 0.8 - \mu(t) \tag{5}$$
$$p(\beta_t = 1|\alpha^t = \alpha^{t-1}0, \beta^{t-1}) = 0.2 + \mu(t) \tag{6}$$
$$p(\beta_t = 0|\alpha^t = \alpha^{t-1}1, \beta^{t-1}) = 0.2 + \mu(t) \tag{7}$$
$$p(\beta_t = 1|\alpha^t = \alpha^{t-1}1, \beta^{t-1}) = 0.8 - \mu(t) \tag{8}$$
$$p(\beta_t = X|\alpha^t = \alpha^{t-1}X, \beta^{t-1}) = 1 \tag{9}$$

[2] If we are interested in transmitting a sequence of bits, the real encoding is a bit more complicated, because we may need to introduce some sort of stop symbol to indicate that the source considers a bit transmitted successfully and it is then proceeding to the transmission of the next bit.

Correspondingly, the channel matrix is:

	0	1	X
$\alpha_t = 0, \beta^{t-1}$	$0.8 + \mu(t)$	$0.2 - \mu(t)$	0
$\alpha_t = 1, \beta^{t-1}$	$0.2 - \mu(t)$	$0.8 + \mu(t)$	0
$\alpha_t = X, \beta^{t-1}$	0	0	1

Now let us consider how to create the code-functions, having in mind that the code-functions depend on the message to be transmited. In time $t = 0$, the code functions are chosen based on the message being transmited. Let us suppose that the message is $W = 0$, the case where $W = 1$ being analougous.

For $t = 1$ corresponds, the channel is used for its first time and the feedback history so far is empty $\beta^0 = \emptyset$. In that case the encoder selects the input symbol $\alpha_0 = 0$ by assigning:

$$\varphi_0[W = 0](\beta^0 = \emptyset) = 0$$

For $t = 2$, there are two possibilities: the feedback history consists of only one bit, and it is either $\beta^1 = 0$ or $\beta^1 = 1$. In the first case, a succesfull transmition occurred, and the encoder can select the end of transmition symbol $\alpha_1 = X$. On the other hand, if $\beta^1 = 1$, some noise occurred during the transmition, and the encoder will try to retransmit the bit. We can write it formally as:

$$\varphi_2[W = 0](\beta^1 = 0) = X$$

$$\varphi_2[W = 0](\beta^1 = 1) = 0$$

In the next round, $t = 3$, the possible feedback histories are $\beta^2 \in \{0X, 10, 11\}$. In the first case, $\beta^2 = 0X$, the presence of the success symbol X indicates that the transmition has already been completed in a correct way, so the encoder just selects X again. If $\beta^2 = 0X$, the transmition had failled in the first try, but has just succeded, so again the encoder just selects X as the new input symbol. In the last case, $\beta^2 = 11$, the transmition has not succeeded yet, so the encoder tries sending the right bit 0 again. Formally:

$$\varphi_3[W = 0](\beta^2 = 0X) = X$$

$$\varphi_3[W = 0](\beta^2 = 10) = X$$

$$\varphi_3[W = 0](\beta^2 = 11) = 0$$

We can generalize the above construction as follows: whenever a bit 1 is fedback, a retransmition of 0 is needed. In the other cases, a succesfull transmition occurred and the encoder considers the transmition completed and selects the X again. Guarding the simetry when the message is $W = 1$, we can write formally, for every $1 \leq t \leq T$:

$W=0$	$W=1$
$\varphi_t[W=0](\beta^{t-1}=1^{t-1})=0$	$\varphi_t[W=1](\beta^{t-1}=0^{t-1})=1$
$\varphi_t[W=0](\beta^{t-1}\neq 1^{t-1})=X$	$\varphi_t[W=1](\beta^{t-1}\neq 1^{t-1})=X$

Once all time steps $0, 1, \ldots, T$ have occurred, the decoder has the whole output history β^T available to try to infer which was the original message W. Our decoder will proceed as follows. By the construction of the encoding scheme, if the received output β^T contains an X, it means that the bit was transmitted correctely, and it is exactly the bit that just preceeds the first occurrence of a symbol X. Otherwise the encoder does not decide whether the intended message was 0 or 1, so it will always assign 0 as the decoded message. Formally:

$$\hat{W} = \gamma(\beta_1, \ldots, 0, X, \ldots, \beta_T) = 0 \tag{10}$$

$$\hat{W} = \gamma(\beta_1, \ldots, 1, X, \ldots, \beta_T) = 1 \tag{11}$$

$$\hat{W} = \gamma(\beta^T) = 0 \qquad \text{in any other case} \tag{12}$$

Table 4 shows a concrete example for the binnary channel with memory and feedback in a scenario where the channel can be used $T = 3$ times and the message being transmitted is $W = 0$.

Time	Code functions	Feedback history	Encoder	Channel	Decoder
$t=0$	Code functions for $W=0$ are selected.	———-	———-	———-	———-
$t=1$	$\varphi_0[W=0](\beta^0=\emptyset)=0$	$\emptyset$	$\alpha_1=\varphi_t(\emptyset)=0$	According to $p(\beta_1\|0,\emptyset)$ produces $\beta_1=1$	———-
$t=2$	$\varphi_0[W=0](\beta^1=0)=X$ $\varphi_0[W=0](\beta^1\neq 0)=0$	1	$\alpha_1=\varphi_t(1)=0$	According to $p(\beta_1\|00,1)$ produces $\beta_1=0$	———-
$t=3$	$\varphi_0[W=0](\beta^2=11)=0$ $\varphi_0[W=0](\beta^2\neq 11)=X$	10	$\alpha_2=\varphi_t(10)=X$	According to $p(\beta_1\|00X,10)$ produces $\beta_1=X$	———-
$t=4$	———-	———-	———-	———-	Decoded message $\hat{W}=\gamma(\beta^3=10X)=0$

Table 4. Evolution of the binary channel with time, for $T = 3$ and $W = 0$

4.2 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 A^T to B^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 information to be transmitted. If we want to maintain the correspondence between the transmission rate and capacity, we need to replace the mutual information with *directed information* [12].

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

In Section 6 we shall discuss relation between directed information and mutual information, as well as the correspondence with information leakage. For the moment, we only present the extension of the concept of capacity.

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 with memory and feedback $\{p(\beta_t|\alpha^t, \beta^{t-1})\}_{t=1}^{T}$ is:

$$C_T = \sup_{\mathcal{D}_T} \frac{1}{T} I(A^T \to B^T) \tag{13}$$

5 Interactive systems as channels with memory and feedback

Interactive Information Hiding Systems (IIHS) [1], 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. Given a general IIHS, it is always possible to find an equivalent one that satisfies this assumptions. The interested reader can find in the appendix the formal definition of the transformation.

Finally, we assume that every state is reachable from the initial state, and 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.

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 $\sigma = \sigma'$.*

Proof. The proof follows easily by induction under the stated assumptions that every state is reachable from the initial state, and that for every state s and label ℓ, there exists a unique state r such that $s \xrightarrow{\ell} r$.

The initial state of the automaton is uniquely determined by the empty (input and output) traces, because every state is reachable. Assume now we are in a state s uniquely determined by input and output traces α and β, respectively. If s makes an input transition $s \xrightarrow{a} s'$, then there is only one state s' reachable from s via an a-transition, and therefore s' is uniquely determined by the input trace $\alpha' = \alpha a$ and the output trace β. Similarly, if s makes an output transition $s \xrightarrow{b} s'$, the state s' is uniquely determined by the input trace α and the output trace $\beta' = \beta b$. □

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 determines a channel as specified by its stochastic kernel, while a fully probabilistic IIHS determines, additionally, also the input distribution. In Section 7 we will give an extensive and detailed example of how to make such a construction for a real security protocol.

Given a path σ of length $2t-1$, we will 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.

The following example shows how to apply Definition 4, with the help of Proposition 1 to build the channel matrix of a simple example.

Example 2. Let us consider an extended version of the website interactive system of Figure 1. We maintain the general definition of the system, i.e, there are two possible buyers ($rich$ and $poor$ represented by $rc.$ and $pr.$, respectively) and two possible products ($cheap$ and $expensive$, represented by $chp.$ and $exp.$, respectively). We still assume that offers are observables, since they are visible to everyone on the website, but the identity of buyers should be kept secret. We consider two consecutive rounds of offers and buys, which implies, after normalization, $T = 3$. Figure 3 shows an automaton for this example in normalized form. Transitions with null probability are omitted.

To construct the channel matrix $\{p(\beta_t|\alpha^t, \beta^{t-1})\}_{t=1}^T$, we need to determine the conditional probability of an observable at time t given the history up to time t.

Let us take the case $t = 2$ and compute the conditional probability of observable $\beta_2 = cheap$ given that the history of secrets until time $t = 2$ is $\alpha^2 = a_*, poor$ and the history of observables is $\beta^1 = expensive$. Applying Definition 4, we see that $p(\beta_2 = cheap|\alpha^2 = a_*, poor, \beta^1 = expensive) = \vartheta(q)(cheap, q')$. By Proposition 1, the traces $\alpha^2 = a_*, poor, \beta^1 = expensive$ determine a unique state q in the automaton, namely, the state $q = 5$. Moreover, from the state 5 a unique transition labelled with the action $cheap$ is possible, leading to the state $q' = 11$. Therefore, we can conclude that $p(\beta_2 = cheap|\alpha^2 = a_*, poor, \beta^1 = expensive) = \vartheta(q = 5)(cheap, q' = 11) = p_{23}$.

Simirlarly, with $t = 1$ and history $\alpha^1 = a_*, \beta^0 = \emptyset$, the output symbol $\beta_1 = expensive$ can be observed with probability $p(\beta_1 = expensive|\alpha^1 = a_*, \beta^0 = \emptyset) = \vartheta(q = 0)(cheap, q' = 2) = \overline{p_1}$.

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}.*

Example 3. Since the system of Example 2 is fully probabilistic, we can calculate the values of the conditional probabilities $\{p(\alpha_t|\alpha^{t-1}, \beta^{t-1})\}_{t=1}^T$.

Let us take as an example the case where $t = 2$ and compute the conditional probability of secret $\alpha_2 = poor$ given that the history of secrets until time $t = 2$ is $\alpha^1 = a_*$ and the history of observables is $\beta^1 = expensive$. Applying Definition 4, we see that $p(\alpha_2 = poor|\alpha_1 = a_*, \beta^1 = expensive) = \vartheta(q)(poor, q')$. By Proposition 1, the traces $\alpha^1 = a_*, \beta^1 = expensive$ determine a unique state q in the automaton, namely, the state $q = 2$. Moreover, from the state 2 a unique transition labelled with the action $poor$ is possible, leading to the state $q' = 5$. Therefore, we can conclude that $p(\alpha_2 = poor|\alpha_1 = a_*, \beta^1 = expensive) = \vartheta(q = 2)(poor, q' = 5) = q_{12}$.

Similarly, with $t = 3$ and history $\alpha^2 = a_*, rich, \beta^2 = cheap, expensive$, the output symbol $\alpha_3 = rich$ can be observed with probability $p(\alpha_3 = rich|\alpha^2 = \alpha_*, rich, \beta^0 = cheap, expensive) = \vartheta(q = 10)(cheap, q' = 21) = \overline{q_{24}}$.

5.1 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. The solution is to externalize this probabilistic behavior to the code functions.

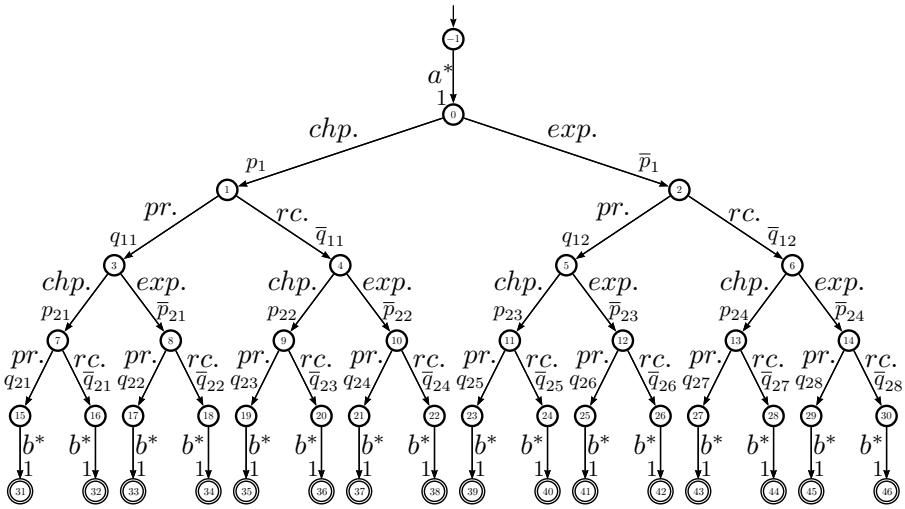


Fig. 3. A longer, normalized version of the interactive system of Figure 1

As showed in [20], the original channel with feedback from input symbols $\mathcal{A}^T$ to output symbols $\mathcal{B}^T$ can be lifted to an equivalent channel without feedback from code functions $\mathcal{F}^T$ to output symbols $\mathcal{B}^T$. This transformation will also allows us to calculate the channel capacity. Let $\{p(\varphi_t|\varphi^{t-1})\}_{t=1}^T$ be a sequence of code function stochastic kernels and let $\{p(\beta_t|\alpha^t,\beta^{t-1})\}_{t=1}^T$ be a channel with memory and feedback. The channel from F^T to B^T is constructed using a joint measure $Q(\varphi^T,\alpha^T,\beta^T)$ that respects the following constraints:

Definition 6. *A measure $Q(\varphi^T,\alpha^T,\beta^T)$ is said to be* consistent *with respect to the code function stochastic kernels $\{p(\varphi_t|\varphi^{t-1})\}_{t=1}^T$ and the channel $\{p(\beta_t|\alpha^t,\beta^{t-1})\}_{t=1}^T$ if, for each t:*

1. *There is no feedback to the code functions: $Q(\varphi_t|\varphi^{t-1},\alpha^{t-1},\beta^{t-1}) = p(\varphi_t|\varphi^{t-1})$.*
2. *The input is a function of the past outputs: $Q(\alpha_t|\varphi^t,\alpha^{t-1},\beta^{t-1}) = \delta_{\{\varphi_t(\beta^{t-1})\}}(\alpha_t)$ where δ is the Dirac measure.*
3. *The properties of the underlying channel are preserved:*

$$Q(\beta_t|F^t=\varphi^t,A^t=\alpha^t,B^{t-1}=\beta^{t-1}) = p(\beta_t|\alpha^t,\beta^{t-1})$$

The following result states that there is only one consistent measure $Q(\varphi^T,\alpha^T,\beta^T)$:

Theorem 4 ([20]). *Given $\{p(\varphi_t|\varphi^{t-1})\}_{t=1}^T$ and a channel $\{p(\beta_t|\alpha^t,\beta^{t-1})\}_{t=1}^T$, there exists only one consistent measure $Q(\varphi^T,\alpha^T,\beta^T)$. Furthermore the channel from $\mathcal{F}^T$ to $\mathcal{B}^T$ is given by:*

$$Q(\beta_t|\varphi^t,\beta^{t-1}) = p(\beta_t|\varphi^t(\beta^{t-1}),\beta^{t-1}) \tag{14}$$

Since in our setting the concept of encoder makes no sense as there is no information to encode, we externalize the probabilistic behavior of α_t as follows. 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). Reaction functions can be seen as a model of how the environment reacts to given system outputs, producing new system inputs (they do not play a role of encoding a message). These reaction functions are endowed with a probability distribution that generates the probabilistic behavior of the values of α_t.

Definition 7. *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.

Lemma 1. *Let* $\mathcal{X}, \mathcal{Y}$ *be finite sets, and let* $\tilde{x} \in \mathcal{X}, \tilde{y} \in \mathcal{Y}$. *Let* $p : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ *be a function such that, for every* $x \in \mathcal{X}$, *we have:* $\sum_{y \in \mathcal{Y}} p(x, y) = 1$. *Then:*

$$\sum_{\substack{f \in \mathcal{X} \to \mathcal{Y} \\ f(\tilde{x}) = \tilde{y}}} \prod_{x \in \mathcal{X}} p(x, f(x)) = p(\tilde{x}, \tilde{y})$$

Proof. By induction on the number of elements of $\mathcal{X}$.

Base case: $X = \{\tilde{x}\}$. In this case:

$$\sum_{\substack{f \in \mathcal{X} \to \mathcal{Y} \\ f(\tilde{x}) = \tilde{y}}} \prod_{x \in \mathcal{X}} p(x, f(x)) = p(\tilde{x}, f(\tilde{x})) = p(\tilde{x}, \tilde{y})$$

Inductive case: Let $\mathcal{X} = \mathcal{X}' \cup \{\mathring{x}\}$, with $\mathring{x} \neq \tilde{x}$, and $\tilde{x} \in \mathcal{X}$. Then:

$$\sum_{\substack{f \in \mathcal{X}' \cup \{\mathring{x}\} \to \mathcal{Y} \\ f(\tilde{x}) = \tilde{y}}} \prod_{x \in \mathcal{X}' \cup \{\mathring{x}\}} p(x, f(x))$$

$$= \quad \text{(by distributivity)}$$

$$\left(\sum_{\substack{f \in \mathcal{X}' \to \mathcal{Y} \\ f(\tilde{x}) = \tilde{y}}} \prod_{x \in \mathcal{X}'} p(x, f(x)) \right) \cdot \sum_{g \in \{\mathring{x}\} \to \mathcal{Y}} p(\mathring{x}, g(\mathring{x}))$$

$=$ (by the assumption)

$$\sum_{\substack{f \in \mathcal{X}' \to \mathcal{Y} \\ f(\tilde{x}) = \tilde{y}}} \prod_{x \in \mathcal{X}'} p(x, f(x))$$

$=$ (by the induction hypothesis)

$p(\tilde{x}, \tilde{y})$

□

Theorem 5. *Given a fully probabilistic* IIHS $\mathcal{I}$*, we can construct a channel with memory and feedback, and probability distribution* $Q(\varphi^T, \alpha^T, \beta^T)$*, which corresponds to* $\mathcal{I}$ *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* $\mathcal{I}$*.*

Proof. First of all we note that, by probability laws, $Q(\alpha^t, \beta^t) = \sum_{\varphi^t} Q(\varphi^t, \alpha^t, \beta^t)$. So we need to show that $\sum_{\varphi^t} Q(\varphi^t, \alpha^t, \beta^t) = p(\alpha^t, \beta^t)$ by induction on t.

Base case: $t = 1$. Let us define $Q(\varphi_1|\epsilon) = p(\varphi_1(\epsilon))$ and $Q(\beta_1|\alpha^1, \epsilon) = p(\beta_1|\alpha_1)$. Then:

$$\begin{aligned}
\sum_{\varphi^1} Q(\varphi^1, \alpha^1, \beta^1) &= \sum_{\varphi_1} Q(\varphi_1, \alpha_1, \beta_1) \\
&= \sum_{\varphi_1} Q(\varphi_1|\epsilon, \epsilon, \epsilon) Q(\alpha_1|\varphi_1, \epsilon, \epsilon) Q(\beta_1|\varphi_1, \alpha_1, \epsilon) && \text{(by the chain rule)} \\
&= \sum_{\varphi_1} Q(\varphi_1|\epsilon) \delta_{\{\varphi_1(\epsilon)\}}(\alpha_1) Q(\beta_1|\alpha^1, \epsilon) && \text{(by Definition 6)} \\
&= \sum_{\varphi_1} p(\varphi_1(\epsilon)) \delta_{\{\varphi_1(\epsilon)\}}(\alpha_1) p(\beta_1|\alpha_1) \\
&= p(\alpha_1) p(\beta_1|\alpha_1) && \text{(by definition of } \delta) \\
&= p(\alpha_1, \beta_1) \\
&= p(\alpha^1, \beta^1)
\end{aligned}$$

Inductive case: Let us define $Q(\beta_t|\alpha^t, \beta^{t-1}) = p(\beta_t|\alpha^t, \beta^{t-1})$, and

$$Q(\varphi_t|\varphi^{t-1}) = \prod_{\beta^{t-1}} p(\varphi_t(\beta^{t-1})|\varphi^{t-1}(\beta^{t-2}), \beta^{t-1})$$

Note that, if we consider $\mathcal{X} = \{\beta^{t-1} \mid \beta_i \in \mathcal{B}, 1 \leq i \leq t-1\}$, $\mathcal{Y} = \mathcal{A}$, and $p(\beta^{t-1}, \alpha_t) = p(\alpha_t|\varphi^{t-1}(\beta^{t-2}), \beta^{t-1})$, then $\mathcal{X}$, $\mathcal{Y}$ and p satisfy the hypothesis of Lemma 1.
Then:

$$\sum_{\varphi^t} Q(\varphi^t, \alpha^t, \beta^t)$$
$=$ (by the chain Rule)
$$\sum_{\varphi^t} Q(\varphi^{t-1}, \alpha^{t-1}, \beta^{t-1}) Q(\varphi_t|\varphi^{t-1}, \alpha^{t-1}, \beta^{t-1}) Q(\alpha_t|\varphi^t, \alpha^{t-1}, \beta^{t-1}) Q(\beta_t|\varphi^t, \alpha^t, \beta^{t-1})$$
$=$ (by Definition 6)
$$\sum_{\varphi^t} Q(\varphi^{t-1}, \alpha^{t-1}, \beta^{t-1}) Q(\varphi_t|\varphi^{t-1}), \delta_{\{\varphi_t(\beta^{t-1})\}}(\alpha_t) Q(\beta_t|\alpha^t, \beta^{t-1})$$
$=$ (by construction of Q)
$$\sum_{\varphi^t} Q(\varphi^{t-1}, \alpha^{t-1}, \beta^{t-1}) \left(\prod_{\beta'^{t-1}} p(\varphi_t(\beta'^{t-1})|\varphi^{t-1}(\beta'^{t-2}), \beta'^{t-1}) \right) \delta_{\{\varphi_t(\beta^{t-1})\}}(\alpha_t) p(\beta_t|\alpha^t, \beta^{t-1})$$
$=$ (by definition of δ)
$$\sum_{\substack{\varphi^t \\ \varphi_t(\beta^{t-1})=\alpha_t}} Q(\varphi^{t-1}, \alpha^{t-1}, \beta^{t-1}) \left(\prod_{\beta'^{t-1}} p(\varphi_t(\beta'^{t-1})|\varphi^{t-1}(\beta'^{t-2}), \beta'^{t-1}) \right) p(\beta_t|\alpha^t, \beta^{t-1})$$
$=$
$$\sum_{\varphi^{t-1}} Q(\varphi^{t-1}, \alpha^{t-1}, \beta^{t-1}) p(\beta_t|\alpha^t, \beta^{t-1}) \sum_{\substack{\varphi_t \\ \varphi_t(\beta^{t-1})=\alpha_t}} \prod_{\beta'^{t-1}} p(\varphi_t(\beta'^{t-1})|\varphi^{t-1}(\beta'^{t-2}), \beta'^{t-1})$$
$=$ (by Lemma 1)
$$\sum_{\varphi^{t-1}} Q(\varphi^{t-1}, \alpha^{t-1}, \beta^{t-1}) \cdot p(\beta_t|\alpha^t, \beta^{t-1}) \cdot p(\alpha_t|\alpha^{t-1}, \beta^{t-1})$$
$=$
$$p(\beta_t|\alpha^t, \beta^{t-1}) \cdot p(\alpha_t|\alpha^{t-1}, \beta^{t-1}) \cdot \sum_{\varphi^{t-1}} Q(\varphi^{t-1}, \alpha^{t-1}, \beta^{t-1})$$
$=$ (by induction hypothesis)
$$p(\beta_t|\alpha^t, \beta^{t-1}) \cdot p(\alpha_t|\alpha^{t-1}, \beta^{t-1}) \cdot p(\alpha^{t-1}, \beta^{t-1})$$
$=$ (by the chain rule)
$$p(\alpha^t, \beta^t)$$

□

Corollary 1. *Let a $\mathcal{I}$ 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 $\mathcal{I}$ according to Definitions 4 and 5. Then the reactor $R = \{p(\varphi_t|\varphi^{t-1})\}_{t=1}^T$ compatible with respect to the $\mathcal{I}$ is given by:*

$$p(\varphi_1) = p(\alpha_1|\alpha^0, \beta^0) = p(\alpha_1) \tag{15}$$

$$p(\varphi_t|\varphi^{t-1}) = \prod_{\beta^{t-1}} p(\varphi_t(\beta^{t-1})|\varphi^{t-1}(\beta^{t-2}), \beta^{t-1}), \quad 2 \leq t \leq T \tag{16}$$

Figure 4 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.

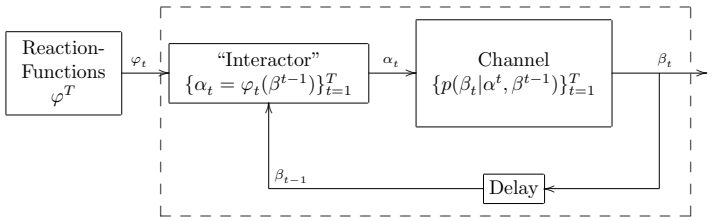


Fig. 4. Channel with memory and feedback model for IIHS

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.

6 Leakage in Interactive Systems

Mutual information expresses *correlation* between two random variables. In fact, the symmetry of mutual information, $I(A^T; B^T) = I(B^T; A^T)$, reflects the fact that the amount of information we can get from B^T by observing A^T is the same amount of information that we get from A^T by observing B^T. There is no concept of *causality*, in the sense that neither A^T is seen as causing B^T, nor the inverse.

Mathematically, for T usages of the channel, the mutual information $I(A^T; B^T)$ can be calculated with the help of the chain law of Equation 2:

$$I(A^T; B^T) = \sum_{t=1}^{T} I(A^T; B_t|B^{t-1}) \tag{17}$$

Note that in the equation above, each term of the sum is the mutual information between the random variable B_t and the whole sequence of random variables $A^T = A_1, \ldots, A_T$, given the history on B^{t-1}. The equation shows that at time $1 \leq t \leq T$, even though only the inputs $\alpha^t = \alpha_1, \alpha_2, \ldots, \alpha_t$ have been fed to the channel, the whole sequence α^T, including $\alpha_{t+1}, \alpha_{t+2}, \ldots, \alpha_T$, has a statistical correlation with the output β_t. Although it can sound surprising at first, we need to remember that in principle the channel can be used with some sort of feedback, and since mutual information is symmetric, it is indeed β_t that has an influence on $\alpha_{t+1}, \alpha_{t+2}, \ldots, \alpha_T$.

Directed information, presented in Definition 1, captures the concept of *causality*, to which the definition of mutual information is indifferent. It splits the correlation between inputs and outputs $I(A^T; B^T)$ into the information that flows from input to output through the channel $I(A^T \rightarrow B^T)$ and the information that flows from output to the input via feedback $I(B^T \rightarrow A^T)$. Note that the directed information is 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 [20] that

$$I(A^T; B^T) = I(A^T \rightarrow B^T) + I(B^T \rightarrow A^T)$$

i.e, the mutual information is the sum of the directed information flow in both senses. That is the reason why it is symmetric.

Once we can split mutual information into directed information in two different directions, it is important to understand the different role that the information flow in each direction plays. The directed information from inputs to outputs $I(A^T \rightarrow B^T)$ represents the system behavior: via the channel the information flows from inputs to outputs according to the system specification, modeled by the channel stochastic kernels. This flow represents the amount of information an attacker can gain from the inputs by observing the outputs and we argue that this is the real information leakage.

On the other hand, the directed information from outputs to inputs $I(A^T \rightarrow B^T)$ represents how the environment reacts to the protocol: given the system outputs, the environment reacts producing new inputs. We argue that the information flow from outputs to inputs induced by this dependence is independent of any particular system, it is a characteristic of the environment itself. Hence, if an attacker knows how the environment reacts to outputs, i.e the probabilistic behavior of the environment reactions given the system outputs, this knowledge is part of the *a priori* knowledge, and should not be counted as leakage.

If a channel does not have feedback, then $I(B^T \rightarrow A^T) = 0$ and it follows that $I(A^T; B^T) = I(A^T \rightarrow B^T)$. In channels without feedback mutual information is a good measure of information flow because it coincides with directed information from input to output. This correspondence does not hold anymore if the channel is used with feedback and, therefore, we should consider the directed information as the real measure of information transmitted by the channel. The following example should help understanding why.

Example 4. 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 5.

	b_1	b_2
a_1	0.5	0.5
a_2	0.5	0.5

Table 5. Channel matrix for Example 4

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 A^t to B^t, since the rows of the matrix are all equal. We have indeed that $I(A^t \to B^t) = 0$, and the mutual information $I(A^t; B^t)$ is only due to the feedback information flow $I(B^t \to A^t)$.

Having in mind the above discussion, we now 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 [3]), 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 4.2), it is natural to consider it as a measure of leakage of information by the protocol.

Definition 8. *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:

$$I(A^T; B^T) = H(A^T) - H(A^T|B^T) \tag{18}$$

The principle behind (18) 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. In the Section 7 we give an extensive and detailed example of how to calculate the leakage for a real security protocol.

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

7 Modeling IIHSs as channels: An example

In this section we show the application of our approach to the *Cocaine Auction Protocol* [18]. 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 formalization 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 $\mathcal{A} = \{\text{Candlemaker}, \text{Scarface}, a^*\}$ 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 5. 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.

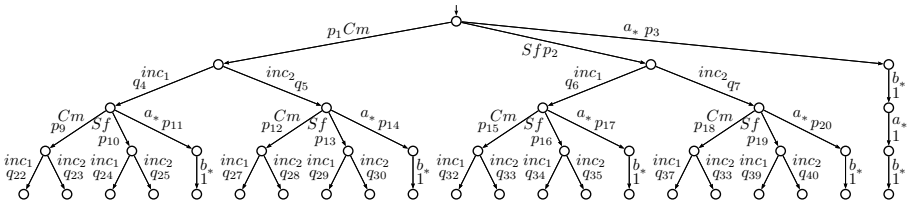


Fig. 5. Cocaine Auction example

Table 6 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[18] suggesting that if the seller knows the identity of the bidders then the (strong) anonymity guaranties are not provided anymore.

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

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

$\alpha_1, \beta_1, \alpha_2 \rightarrow \beta_2$	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 6. Stochastic kernels for the Cocaine Auction example.

The next step is to construct all the possible reaction functions $\{f_t(\beta^{t-1})\}_{t=1}^{T}$. As seen in Section 5.1, the reaction functions are the correspondent to the encoder in the channel. They take the feedback story and decide how the world is going to react to this situation. For this example, Table 7 shows the reaction functions for each time t.

Now we need to define the reactor, i.e., the reaction functions stochastic kernel. Corollary 1 shows that we can do so by using the following equations:

$$p(\varphi_1) = p(\alpha_1|\alpha^0, \beta^0) = p(\alpha_1)$$
$$p(\varphi_t|\varphi^{t-1}) = \prod_{\beta^{t-1}} p(\varphi_t(\beta^{t-1})|\varphi^{t-1}(\beta^{t-2}), \beta^{t-1}), \quad 2 \leq t \leq T$$

For instance, $p(f_{1(1)}) = p(Candlemaker) = p_1$. In the same way, $p(f_{1(2)}) = p(\textit{Scarface}) = p_2$ and $p(f_{1(3)}) = p(a_*) = p_3$.

Let us take as an example the calculation of $p(f_{2(6)}|f_{1(3)})$:

β^0	$f_{1(1)}$	$f_{1(2)}$	$f_{1(3)}$
$\emptyset$	*Candlemaker*	*Scarface*	a_*

(a) All 3 reaction functions φ_1

β^1	$f_{2(1)}(\beta^1)$	$f_{2(2)}(\beta^1)$	$f_{2(3)}(\beta^1)$	$f_{2(4)}(\beta^1)$	$f_{2(5)}(\beta^1)$	$f_{2(6)}(\beta^1)$	$f_{2(7)}(\beta^1)$
inc_1	*Candlemaker*	*Candlemaker*	*Candlemaker*	*Candlemaker*	*Candlemaker*	*Candlemaker*	*Candlemaker*
inc_2	*Candlemaker*	*Candlemaker*	*Candlemaker*	*Scarface*	*Scarface*	*Scarface*	a_*
b_*	*Candlemaker*	*Scarface*	a_*	*Candlemaker*	*Scarface*	a_*	*Candlemaker*
β^1	$f_{2(8)}(\beta^1)$	$f_{2(9)}(\beta^1)$	$f_{2(10)}(\beta^1)$	$f_{2(11)}(\beta^1)$	$f_{2(12)}(\beta^1)$	$f_{2(13)}(\beta^1)$	$f_{2(14)}(\beta^1)$
inc_1	*Candlemaker*	*Candlemaker*	*Scarface*	*Scarface*	*Scarface*	*Scarface*	*Scarface*
inc_2	a_*	a_*	*Candlemaker*	*Candlemaker*	*Candlemaker*	*Scarface*	*Scarface*
b_*	*Scarface*	a_*	*Candlemaker*	*Scarface*	a_*	*Candlemaker*	*Scarface*
β^1	$f_{2(15)}(\beta^1)$	$f_{2(16)}(\beta^1)$	$f_{2(17)}(\beta^1)$	$f_{2(18)}(\beta^1)$	$f_{2(19)}(\beta^1)$	$f_{2(20)}(\beta^1)$	$f_{2(21)}(\beta^1)$
inc_1	*Scarface*	*Scarface*	*Scarface*	*Scarface*	a_*	a_*	a_*
inc_2	*Scarface*	a_*	a_*	a_*	*Candlemaker*	*Candlemaker*	*Candlemaker*
b_*	a_*	*Candlemaker*	*Scarface*	a_*	*Candlemaker*	*Scarface*	a_*
β^1	$f_{2(22)}(\beta^1)$	$f_{2(23)}(\beta^1)$	$f_{2(24)}(\beta^1)$	$f_{2(25)}(\beta^1)$	$f_{2(26)}(\beta^1)$	$f_{2(27)}(\beta^1)$	-
inc_1	a_*	a_*	a_*	a_*	a_*	a_*	-
inc_2	*Scarface*	*Scarface*	*Scarface*	a_*	a_*	a_*	-
b_*	*Candlemaker*	*Scarface*	a_*	*Candlemaker*	*Scarface*	a_*	-

(b) All 27 reaction functions $\varphi_2(\beta^1)$

Table 7. Reaction functions for the cocaine auction example.

$$
\begin{aligned}
p(f_{2(6)}|f_{1(1)}) &= \prod_{\beta^1} p(f_{2(6)}(\beta^1)|\varphi_{1(1)}, \beta^1) \\
&= p(f_{2(6)}(inc_1)|\textit{Candlemaker}, inc_1) \cdot p(f_{2(6)}(inc_2)|\textit{Candlemaker}, inc_2) \\
&\quad p(f_{2(6)}(b_*)|\textit{Candlemaker}, b_*) \\
&= p(\textit{Candlemaker}|\textit{Candlemaker}, inc_1) \cdot p(\textit{Scarface}|\textit{Candlemaker}, inc_2) \\
&\quad p(a_*|\textit{Candlemaker}, b_*) \\
&= p_9 \cdot p_{13} \cdot 1 \\
&= p_9 p_{13} \qquad (19)
\end{aligned}
$$

Note that some reaction functions can have probability 0, which is consistent with probabilistic automaton. For instance:

$$\begin{aligned} p(f_{2(25)}|f_{1(3)}) &= \prod_{\beta^1} p(f_{2(4)}(\beta^1)|\varphi_{1(3)}, \beta^1) \\ &= p(f_{2(4)}(inc_1)|a_*, inc_1) \cdot p(f_{2(4)}(inc_2)|a_*, inc_2) \cdot p(f_{2(4)}(b_*)|a_*, b_*) \\ &= p(b_*|a_*, inc_1) \cdot p(b_*|a_*, inc_2) \cdot p(\textit{Candlemaker}|a_*, b_*) \\ &= 1 \cdot 1 \cdot 0 \\ &= 0 \end{aligned} \tag{20}$$

7.1 Calculating the information leakage

Let us now calculate the information leakage for this cocaine auction example using the concepts from Section 6. We are going to analyze three different scenarios:

Example a: There is feedback, but the probability of an observable does not depend on the history of secrets. In the auction protocol, this corresponds to a scenario where the probability of one of the mob members to bid can depend on the increment imposed by the seller, but the history of who has previously bid in the past has no influence on the choice of increments by the seller during the coming turns. In other words, the seller cannot use the information of who has been bidding to change his strategy of defining the new increments. This situation corresponds to the original description of the protocol in [18], where the seller does not have access to the identity of the bidder, for the sake of anonymity preservation. In general, we have that $p(\beta_t|\alpha^t, \beta^{t-1}) = p(\beta_t|\beta^{t-1})$ for every $1 \leq t \leq T$. However, there is an exception: if there is no bidder, case modeled by the secret being a_*, then the auction terminates, which is signaled by the observable b_*.

Example b: This is most general case: no restrictions. The presence of feedback allows the probability of the bidder to depend on the increment on the price. For instance, if *Candlemaker* is richer than *Scarface*, it is more likely that the latter bids if the increment in the price is inc_2 instead of inc_1. Also, the probability of an observable can depend on the history of secrets, i.e., in general $p(\beta_t|\alpha^t, \beta^{t-1}) \neq p(\beta_t|\beta^{t-1})$ for $1 \leq t \leq T$. This scenario can represent a situation where the seller is corrupted and can use his information to affect the outcome of the auction. As an example, suppose that the seller is a friend of *Scarface* and he wants to help him in the auction. One way of doing so is to check who was the winner of the last bidding round. Whenever the winner is *Candlemaker*, the seller chooses for increment the small value inc_1, hoping that it will give *Scarface* a good chance to bid in the next round. On the other hand, whenever the seller detects that the winner is *Scarface*, he chooses for the next increment the greater value inc_2, hoping that it will minimize the chances of *Candlemaker* to bid in the next round (and therefore maximizing the chances of the auction to end having *Scarface* as the final winner).

Example c: There is no feedback. In the cocaine auction, we can have the (maybe unrealistic) situation in which the increment added to the bid has no influence on

the probability of *Candlemaker* or *Scarface* being the bidder. Mathematically, we have that $p(\alpha_t|\alpha^{t-1}, \beta^{t-1}) = p(\alpha_t|\alpha^{t-1})$ for every $1 \leq t \leq T$. However, like in Example b, we do not impose any restriction to $p(\beta_t|\alpha^t, \beta^{t-1})$.

For each scenario we need to attribute values to the probabilities in the protocol tree in Figure 5. The probabilities for each example are listed in Table 8.

Probability variable	Example a value	Example b value	Example c value
p_1	0.7	0.7	0.7
p_2	0.2	0.2	0.2
p_3	0.1	0.1	0.1
q_4	0.9	0.1	0.1
q_5	0.1	0.9	0.9
q_6	0.9	0.9	0.9
q_7	0.1	0.1	0.1
p_9	0.6	0.6	0.6
p_{10}	0.3	0.3	0.3
p_{11}	0.1	0.1	0.1
p_{12}	0.5	0.5	0.6
p_{13}	0.3	0.3	0.3
p_{14}	0.2	0.2	0.1
p_{15}	0.4	0.4	0.5
p_{16}	0.4	0.4	0.2
p_{17}	0.2	0.2	0.3
p_{18}	0.6	0.6	0.5
p_{19}	0.3	0.3	0.2
p_{20}	0.1	0.1	0.3
q_{22}	0.4	0.1	0.1
q_{23}	0.6	0.9	0.9
q_{24}	0.7	0.9	0.9
q_{25}	0.3	0.1	0.1
q_{27}	0.2	0.1	0.1
q_{28}	0.8	0.9	0.9
q_{29}	0.1	0.9	0.9
q_{30}	0.9	0.1	0.1
q_{32}	0.4	0.1	0.1
q_{33}	0.6	0.9	0.9
q_{34}	0.7	0.9	0.9
q_{35}	0.3	0.1	0.1
q_{37}	0.2	0.1	0.1
q_{38}	0.8	0.9	0.9
q_{39}	0.1	0.9	0.9
q_{40}	0.9	0.1	0.1

Table 8. Values of the probabilities in Figure 5 in 3 different examples.

Table 9 shows a comparison between some relevant values on the three cases.

Interpretation	Symbol	Example a	Example b	Example c
Input uncertainty	$H(A^T)$	2.3833	2.4891	2.3607
Reactor uncertainty	H_R	2.3768	2.4832	2.3607
A posteriori uncertainty	$H(A^T\|B^T)$	1.3683	0.0677	0.6646
Mutual information	$I(A^T;B^T) = H(A^T) - H(A^T\|B^T)$	1.0150	1.8214	1.6961
Leakage	$I(A^T \to B^T) = H_R - H(A^T\|B^T)$	1.0085	1.8155	1.6961
Feedback information	$I(B^T \to A^T)$	0.185955	0.0060	0.0000

Table 9. Values for the examples.

In Example a, since the probability of observables does not depend on the history of secrets, there is (almost) no information flowing from the input to the output, and the directed information $I(A^T \to B^T)$ is close to zero, i.e., there leakage is low. The only reason why the leakage is not zero is because the end of an auction needs to be signaled. However, due to presence of feedback, the directed information in the other sense $I(B^T \to A^T)$ is non-zero, and so is the mutual information $I(A^T;B^T)$. This is an example where the mutual information does not correspond to the real information leakage, since some (in this case, most) of the correlation between input and output can be attributed to the feedback.

In Example b the information flow from input to output $I(A^T \to B^T)$ is significantly higher than zero, but still, due to feedback, the information flow from outputs to inputs $I(B^T \to A^T)$ is not zero and the mutual information $I(A^T;B^T)$ is higher than the directed information $I(A^T \to B^T)$ which gives the actual leakage.

In Example c, the absence of feedback implies that $I(B^T \to A^T)$ is zero. In that case the values of $I(A^T;B^T)$ and $I(A^T \to B^T)$ coincide, and correspond to leakage.

8 Topological properties of IIHSs and their Capacity

In this section we show how to extend to IIHSs the notion of pseudometric defined in [7] 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.

Note that, in contrast to metrics, in pseudometrics two elements can have distance 0 without being identical. The reason for considering pseudometrics instead than metrics is because the purpose is to extend the notion of (probabilistic) bisimulation: having distance 0 will correspond to being bisimilar.

We now define a complete lattice on pseudometrics, in order to define the distance between IIHSs as the greatest fixpoint of a particular transformation, in line with the coinductive theory of bisimilarity. Since larger bisimulations identify more, the natural extension of the ordering to metrics must shorten the distances as we go up in the lattice:

Definition 10. *$\mathcal{M}$ is the class of 1-bounded pseudometrics on states with the ordering*

$$d \preceq d' \text{ if } \forall s, s' \in S : d(s, s') \geq d'(s, s').$$

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 [21, 7, 6], we apply the construction based on the Kantorovich metric [10].

Definition 11. *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*

$$\hat{d}((\ell_i, s_i), (\ell_j, s_j)) = \begin{cases} 1 & \text{if } \ell_i \neq \ell_j \\ d(s_i, s_j) & \text{otherwise} \end{cases}$$

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

Definition 12. *$d \in \mathcal{M}$ 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 $d(\mu, \mu') \leq \epsilon$.*

Note that it is not necessary to require the converse of the condition in Definition 12 to get a complete analogy with bisimulation: the converse is indeed implied by the symmetry of d as a pseudometric. Note also that we prohibit ϵ to be 1 because throughout this paper 1 represents the maximum distance, which includes the case where one state may perform a transition and the other may not.

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 $\mathcal{M}$. Eventually we are interested in the distance between IIHSs, and for the sake of simplicity, from now on we consider only the distance between states belonging to different IIHSs. The extension to the general case is trivial. For clarity purposes, we assume that different IIHSs have disjoint sets of states.

Definition 13. *Given two* IIHSs *with transition relations θ and θ' respectively, and a preudometric d on states, define $\Phi : \mathcal{M} \to \mathcal{M}$ as:*

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

It is easy to see that the definition of Φ is a particular case of the function F defined in [7, 6]. Hence it can be proved, by adapting the proofs of the analogous results in [7, 6], that $F(d)$ is a pseudometric, and that the following property holds.

Lemma 2. *For $\epsilon \in [0, 1)$, $\Phi(d)(s, s') \leq \epsilon$ holds if and only if whenever $s \to \mu$, there exists some μ' such that $s' \to \mu'$ and $d(\mu, \mu') \leq \epsilon$.*

Corollary 2. *d is a bisimulation metric iff $d \preceq \Phi(d)$.*

As a consequence of Corollary 2, we have that $d_{max} = \bigsqcup\{d \in \mathcal{M} \mid d \preceq \Phi(d)\}$, and still as a particular case of F in [7, 6], we have that Φ is monotonic on $\mathcal{M}$.

We can now apply Tarski's fixed point theorem, which ensures that d_{max} is the greatest fixed point of Φ. Furthermore, by Corollary 2 we know 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 [7], available on the authors' web pages). Therefore one can proceed in a standard way to 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 $d(\mathcal{I}, \mathcal{I}') = d_{max}(s, s')$. The following properties are auxiliary to the theorem which states the continuity of the capacity.

Lemma 3. *Consider two* IIHSs *$\mathcal{I}$ and $\mathcal{I}'$ with transition functions ϑ and ϑ' respectively. Given $t \geq 2$ and two sequences α^t and β^t, assume that both $\mathcal{I}(\alpha^{t-1}, \beta^{t-1})$ and $\mathcal{I}'(\alpha^{t-1}, \beta^{t-1})$ are defined, that $d_{max}(\mathcal{I}(\alpha^{t-1}, \beta^{t-1}), \mathcal{I}'(\alpha^{t-1}, \beta^{t-1})) < p(\beta_t \mid \alpha^t, \beta^{t-1})$, and $\vartheta(\mathcal{I}(\alpha^t, \beta^{t-1})) \neq \emptyset$. Then:*

1. *$\vartheta'(\mathcal{I}'(\alpha^t, \beta^{t-1})) \neq \emptyset$ holds as well,*
2. *$\mathcal{I}(\alpha^t, \beta^t)$ and $\mathcal{I}'(\alpha^t, \beta^t)$ are both defined, $p(\beta_t \mid \alpha^t, \beta^{t-1}) > 0$, and*

$$d_{max}(\mathcal{I}(\alpha^t, \beta^t), \mathcal{I}'(\alpha^t, \beta^t)) \leq \frac{d_{max}(\mathcal{I}(\alpha^{t-1}, \beta^{t-1}), \mathcal{I}'(\alpha^{t-1}, \beta^{t-1}))}{p(\beta_t \mid \alpha^t, \beta^{t-1}).}$$

Proof.

1. Assume $\vartheta(\mathcal{I}(\alpha^t, \beta^{t-1})) \neq \emptyset$ and, by contradiction, $\vartheta'(\mathcal{I}'(\alpha^t, \beta^{t-1})) = \emptyset$. Since d_{max} is a fixed point of F, we have $d_{max} = F(d_{max})$, and therefore

$$\begin{aligned} d_{max}(\mathcal{I}(\alpha^t, \beta^{t-1}), \mathcal{I}'(\alpha^t, \beta^{t-1})) &= F(d_{max})(\mathcal{I}(\alpha^t, \beta^{t-1}), \mathcal{I}'(\alpha^t, \beta^{t-1})) \\ &= 1 \\ &\geq p(\beta_t \mid \alpha^t, \beta^{t-1}), \end{aligned}$$

   against the hypothesis.

2. If $\vartheta(\mathcal{I}(\alpha^t, \beta^{t-1})) \neq \emptyset$, then, by the first point of this lemma, $\vartheta'(\mathcal{I}'(\alpha^t, \beta^{t-1})) \neq \emptyset$ holds as well, and therefore both $\mathcal{I}(\alpha^t, \beta^t)$ and $\mathcal{I}'(\alpha^t, \beta^t)$ are defined. The hypothesis $d_{max}(\mathcal{I}(\alpha^{t-1}, \beta^{t-1}), \mathcal{I}'(\alpha^{t-1}, \beta^{t-1})) < p(\beta_t \mid \alpha^t, \beta^{t-1})$ ensures that

$p(\beta_t \mid \alpha^t, \beta^{t-1}) < 0$. Let us now prove the bound on $d_{max}(\mathcal{I}(\alpha^t, \beta^t), \mathcal{I}'(\alpha^t, \beta^t))$. By definition of Φ, we have

$$\Phi(d_{max})(\mathcal{I}(\alpha^{t-1}, \beta^{t-1}), \mathcal{I}'(\alpha^{t-1}, \beta^{t-1})) \geq d_{max}(\mathcal{I}(\alpha^t, \beta^{t-1}), \mathcal{I}'(\alpha^t, \beta^{t-1})).$$

Since $d_{max} = \Phi(d_{max})$, we have

$$d_{max}(\mathcal{I}(\alpha^{t-1}, \beta^{t-1}), \mathcal{I}'(\alpha^{t-1}, \beta^{t-1})) \geq d_{max}(\mathcal{I}(\alpha^t, \beta^{t-1}), \mathcal{I}'(\alpha^t, \beta^{t-1})). \quad (21)$$

By definition of Φ and of the Kantorovich metric, we have

$$\begin{aligned}\Phi(d_{max})(\mathcal{I}(\alpha^t, \beta^{t-1}), \mathcal{I}'(\alpha^t, \beta^{t-1})) \geq\ & p(\beta_t \mid \alpha^t, \beta^{t-1}) \cdot \\ & d_{max}(\mathcal{I}(\alpha^t, \beta^t), \mathcal{I}'(\alpha^t, \beta^t)).\end{aligned}$$

Using again $d_{max} = \Phi(d_{max})$, we get

$$\begin{aligned}d_{max}(\mathcal{I}(\alpha^t, \beta^{t-1}), \mathcal{I}'(\alpha^t, \beta^{t-1})) \geq\ & p(\beta_t \mid \alpha^t, \beta^{t-1}) \cdot \\ & d_{max}(\mathcal{I}(\alpha^t, \beta^t), \mathcal{I}'(\alpha^t, \beta^t)),\end{aligned}$$

which, together with (21), allows us to conclude.

□

Lemma 4. *Consider two* IIHSs $\mathcal{I}$ *and* $\mathcal{I}'$*, and let* $p(\cdot \mid \cdot, \cdot)$ *and* $p'(\cdot \mid \cdot, \cdot)$ *be their distributions on the output nodes. Given* $T > 0$*, and two sequences* α^T *and* β^T*, assume that* $p(\beta_t \mid \alpha^t, \beta^{t-1}) > 0$ *for every* $t < T$*. Let* $m = \min_{1 \leq t < T} p(\beta_t \mid \alpha^t, \beta^{t-1})$ *and let* $\epsilon \in (0, m^{T-1})$*. Assume* $d(\mathcal{I}, \mathcal{I}') < \epsilon$*. Then, for every* $t \leq T$*, we have*

$$p(\beta_t \mid \alpha^t, \beta^{t-1}) - p'(\beta_t \mid \alpha^t, \beta^{t-1}) < \frac{\epsilon}{m^{T-1}}.$$

Proof. Observe that, for every $t < T$, $\mathcal{I}(\alpha^t, \beta^t)$ must be defined, and, by repeatedly applying Lemma 3(1), we get that also $\mathcal{I}'(\alpha^t, \beta^t)$ is defined. By definition of φ, and of the Kantorovich metric, we have

$$p(\beta_t \mid \alpha^t, \beta^{t-1}) - p'(\beta_t \mid \alpha^t, \beta^{t-1}) \leq \Phi(d_{max})(\mathcal{I}(\alpha^{t-1}, \beta^{t-1}), \mathcal{I}'(\alpha^{t-1}, \beta^{t-1})),$$

and since d_{max} is a fixed point of Φ, we get

$$p(\beta_t \mid \alpha^t, \beta^{t-1}) - p'(\beta_t \mid \alpha^t, \beta^{t-1}) \leq d_{max}(\mathcal{I}(\alpha^{t-1}, \beta^{t-1}), \mathcal{I}'(\alpha^{t-1}, \beta^{t-1})). \quad (22)$$

By applying $t - 1$ times Lemma 3(2), from (22) we get

$$\begin{aligned}p(\beta_t \mid \alpha^t, \beta^{t-1}) - p'(\beta_t \mid \alpha^t, \beta^{t-1}) &\leq \tfrac{d_{max}(\mathcal{I}(\alpha^0, \beta^0), \mathcal{I}'(\alpha^0, \beta^0))}{m^{t-1}} \\ &= \tfrac{d(\mathcal{I}, \mathcal{I}')}{m^{t-1}} \\ &\leq \tfrac{d(\mathcal{I}, \mathcal{I}')}{m^{T-1}} \\ &< \tfrac{\epsilon}{m^{T-1}}\end{aligned}$$

□

Note that previous lemma states a sort of continuity property of the matrices obtained from IIHSs, but not uniform continuity, because of the dependence on one of the two IIHSs. It is easy to see (from the proof of the Lemma) that uniform continuity does not hold.

The main contribution of this section, stated in next theorem, is the continuity of the capacity w.r.t. the metric on IIHSs. For this theorem, we assume that the IIHSs are normalized. Furthermore, it is crucial that they are secret-nondeterministic (while the definition of the metric holds in general).

Theorem 6. *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* $|C_T(\mathcal{I}) - C_T(\mathcal{I}')| < \epsilon$.

Proof. Consider two normalized IIHSs $\mathcal{I}$ and $\mathcal{I}'$ and choose $T, \epsilon > 0$. Observe that

$$\begin{aligned} |C_T(\mathcal{I}) - C_T(\mathcal{I}')| &= |\max_{p_F(\cdot)} \frac{1}{T} I(A^T \to B^T) - \max_{p_F(\cdot)} \frac{1}{T} I(A'^T \to B'^T)| \\ &\leq \tfrac{1}{T} \max_{p_F(\cdot)} |I(A^T \to B^T) - I(A'^T \to B'^T)| \end{aligned}$$

Since the directed information $I(A^T \to B^T)$ is defined by means of arithmetic operations and logarithms on the joint probabilities $p(\alpha^t, \beta^t)$ and on the conditional probabilities $p(\alpha^t, \beta^t)$, $p(\alpha^t, \beta^{t-1})$, which in turn can be obtained by means of arithmetic operations from the probabilities $p(\beta_t \mid \alpha^t, \beta^{t-1})$ and $p_F(\varphi^t)$, we have that $I(A^T \to B^T)$ is a continuous functions of the distributions $p(\beta_t \mid \alpha^t, \beta^{t-1})$ and $p_F(\varphi^t)$, for every $t \leq T$. Let $p(\beta_t \mid \alpha^t, \beta^{t-1})$, $p'(\beta_t \mid \alpha^t, \beta^{t-1})$ be the distributions on the output nodes of $\mathcal{I}$ and $\mathcal{I}'$, modified in the following way: starting from level T, whenever $p(\beta_t \mid \alpha^t, \beta^{t-1}) = 0$, then we redefine the distributions in all the output nodes of the subtree rooted in $\mathcal{I}(\alpha^t, \beta^t)$ so that they coincide with the distribution of the corresponding nodes of in $\mathcal{I}'$, and analogously for $p'(\beta_t \mid \alpha^t, \beta^{t-1})$. Note that this transformation does not change the directed information, because the subtree rooted in $\mathcal{I}(\alpha^t, \beta^t)$ does not contribute to it, due to the fact that it depends the probability of reaching any of its nodes is 0. The continuity of $I(A^T \to B^T)$ implies that there exists $\epsilon' > 0$ such that, if $|p(\beta_t \mid \alpha^t, \beta^{t-1}) - p'(\beta_t \mid \alpha^t, \beta^{t-1})| < \epsilon'$ for all $t \leq T$ and all sequences α^t, β^t, then, for any $p_F(\varphi^t)$, we have $|I(A^T \to B^T) - I(A'^T \to B'^T)| < \epsilon$. The result then follows from Lemma 4, by choosing

$$\nu = \epsilon' \cdot \min\Big(\min_{\substack{1 \leq t < T \\ p(\beta_t \mid \alpha^t, \beta^{t-1}) > 0}} p(\beta_t \mid \alpha^t, \beta^{t-1}), \min_{\substack{1 \leq t < T \\ p'(\beta_t \mid \alpha^t, \beta^{t-1}) > 0}} p'(\beta_t \mid \alpha^t, \beta^{t-1}) \Big).$$

□

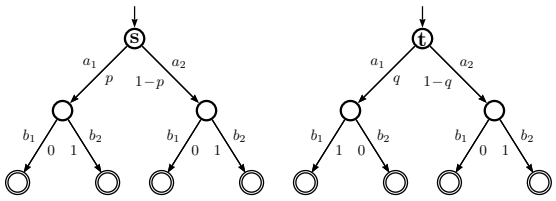
We conclude this section with an example showing that the continuity result for the capacity does not hold if the construction of the channel is done starting from a system in which the secrets are endowed with a probability distribution. This is also the reason why we could not simply adopt the proof technique of the continuity result in [7] and we had to come up with a different reasoning.

Example 5. 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) $(send(a_1) +_q\ send(a_2)) \parallel receive(x).if\ x = a_1\ then\ output(b_1)\ else\ output(b_2).$

Table 10 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 10. The IIHSs of Example 5 and their corresponding channels

9 Conclusion and discussion

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 11(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.

Thorough the paper we have assumed Principle 3. What happens if this assumption is removed? First f all, we observe that the removal could make sense, i.e. in certain cases we could argue that the probabilistic knowledge associated to the secret choices (and its dependence on the observables) *could be considered as part of the leakage*. In the cases a and b of the cocaine auction example in Section 7, for instance, one may want to consider the information that we can deduce about the secrets (the identities of the bidder) from the observables (the increments of the seller) as a leak due to the protocol. Our framework can encompass also this case, and the model remains the same. But the leakage would be represented by the mutual information rather than by the directed one.

IIHSs **as automata**	IIHSs **as channels**	**Notion of leakage**
Normalized IIHSs with nondeterministic inputs and probabilistic outputs	Sequence of stochastic kernels $\{p(\beta_t \mid \alpha^t, \beta^{t-1})\}_{t=1}^{T}$	Leakage as capacity
Normalized IIHSs with a deterministic scheduler solving the nondeterminism	Sequence of stochastic kernels $\{p(\beta_t \mid \alpha^t, \beta^{t-1})\}_{t=1}^{T}$ + reaction function seq. φ^T	
Fully probabilistic normalized IIHSs	Sequence of stochastic kernels $\{p(\beta_t \mid \alpha^t, \beta^{t-1})\}_{t=1}^{T}$ + reactor $\{p(\varphi_t \mid \varphi^{t-1})\}_{t=1}^{T}$	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 11.

In some other cases the flow of information from the observables to the secrets may even be considered as a consequence of the active attacks of an adversary, which uses the observables to modify the probability of the secrets. In this case the leakage would be divided in two parts: the one due to the protocol, represented by $I(A^T \to B^T)$, and the one due to the attacks of the adversaries, and represented by $I(B^T \to A^T)$. THe total leakage would still be represented by the mutual information.

10 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.

We intend to study the applicability of our framework to the area of game theory. In particular, the interactive nature of games such as *Prisoner Dilemma* [14] and *Stag and Hunt* [16] (in their iterative versions) can be modeled as channels with memory and feedback following the techniques proposed in this work. Furthermore, (probabilistic) strategies can be encoded as reaction functions. In this way, optimal strategies are attained by reaction functions maximizing the leakage of the channel.

Acknowledgement

We wish to thank the anonymous reviewers of CONCUR 2010 for their useful comments and recommendations.

References

1. M. E. Andrés, C. Palamidessi, P. van Rossum, and G. Smith. Computing the leakage of information-hiding systems. In J. Esparza and R. Majumdar, editors, *Proceedings of the Sixteenth International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, volume 6015 of *Lecture Notes in Computer Science*, pages 373–389. Springer, 2010.
2. A. Bohannon, B. C. Pierce, V. Sjöberg, S. Weirich, and S. Zdancewic. Reactive noninterference. In E. Al-Shaer, S. Jha, and A. D. Keromytis, editors, *Proceedings of the 2009 ACM Conference on Computer and Communications Security, CCS 2009, Chicago, Illinois, USA, November 9-13, 2009*, pages 79–90. ACM, 2009.
3. K. Chatzikokolakis, C. Palamidessi, and P. Panangaden. Anonymity protocols as noisy channels. *Inf. and Comp.*, 206(2–4):378–401, 2008.
4. D. Clark, S. Hunt, and P. Malacaria. Quantified interference for a while language. In *Proceedings of the Second Workshop on Quantitative Aspects of Programming Languages (QAPL 2004)*, volume 112 of *Electronic Notes in Theoretical Computer Science*, pages 149–166. Elsevier Science B.V., 2005.
5. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., second edition, 2006.
6. Y. Deng, T. Chothia, C. Palamidessi, and J. Pang. Metrics for action-labelled quantitative transition systems. In *Proceedings of the Third Workshop on Quantitative Aspects of Programming Languages (QAPL 2005)*, volume 153 of *Electronic Notes in Theoretical Computer Science*, pages 79–96. Elsevier Science Publishers, 2006. `http://www.lix.polytechnique.fr/~catuscia/papers/Metrics/QAPL/gts.pdf`.
7. J. Desharnais, R. Jagadeesan, V. Gupta, and P. Panangaden. The metric analogue of weak bisimulation for probabilistic processes. In *Proceedings of the 17th Annual IEEE Symposium on Logic in Computer Science*, pages 413–422. IEEE Computer Society, 2002.
8. Ebay website. `http://www.ebay.com/`.
9. Ebid website. `http://www.ebid.net/`.
10. 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.

11. P. Malacaria. Assessing security threats of looping constructs. In M. Hofmann and M. Felleisen, editors, *Proceedings of the 34th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2007, Nice, France, January 17-19, 2007*, pages 225–235. ACM, 2007.
12. J. L. Massey. Causality, feedback and directed information. In *Proc. of the 1990 Intl. Symposium on Information Theory and its Applications*, November 1990.
13. Mercadolibre website. `http://www.mercadolibre.com/`.
14. W. Poundstone. *Prisoners Dilemma*. Doubleday NY, 1992.
15. R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, June 1995. Tech. Rep. MIT/LCS/TR-676.
16. B. Skyrms. *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, 2003.
17. G. Smith. On the foundations of quantitative information flow. In L. de Alfaro, editor, *Proc. of the 12th Int. Conf. on Foundations of Software Science and Computation Structures*, volume 5504 of *LNCS*, pages 288–302, York, UK, 2009. Springer.
18. F. Stajano and R. J. Anderson. The cocaine auction protocol: On the power of anonymous broadcast. In *Information Hiding*, pages 434–447, 1999.
19. S. Subramanian. Design and verification of a secure electronic auction protocol. In *Proceedings of the 17th IEEE Symposium on Reliable Distributed Systems*, pages 204–210, Los Alamitos, CA, USA, 1998. IEEE Computer Society.
20. S. Tatikonda and S. K. Mitter. The capacity of channels with feedback. *IEEE Transactions on Information Theory*, 55(1):323–349, 2009.
21. F. van Breugel and J. Worrell. Towards quantitative verification of probabilistic transition systems. In F. Orejas, P. G. Spirakis, and J. van Leeuwen, editors, *Proceedings of the 28th International Colloquium on Automata, Languages and Programming (ICALP)*, volume 2076 of *Lecture Notes in Computer Science*, pages 421–432. Springer, 2001.
22. W. Vickrey. Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance*, 16(1):8–37, 1961.

Appendix

A: Normalization of IIHS trees

In this section we will address the problem of *normalizing* an IIHS in such a way it is compatible with the assumptions made along the paper. The process of normalization described bellow is general enough to be applied to any IIHS without loss of generality or expression power.

Consider a general IIHS $\mathcal{I} = (M, \mathcal{A}, \mathcal{B}, \mathcal{L}_\tau)$ with $M = (Q, \mathcal{L}, \hat{s}, \vartheta)$, where $\mathcal{L} = \mathcal{A} \cup \mathcal{B} \cup \mathcal{L}_\tau$. Let $\mathcal{I}$ represent an interactive system, such as a protocol. Let us consider that we are interested only in a finite execution of the protocol, so the automaton tree is already unfolded up to a certain level in such a way that the longest input trace is $\alpha^{T'}$ and the longest output trace is $\beta^{T''}$.

In the normalization process, first we will extend the input alphabet $\mathcal{A}$ with a new symbol $a_* \notin \mathcal{A}$ and, in the same way, the output alphabet $\mathcal{B}$ with a new symbol $b_* \notin \mathcal{B}$. As we will see soon, those new symbols will be used as placeholders when we need to rebalance the tree. The new input and output alphabets will be referred as $\mathcal{A}' = \mathcal{A} \cup \{a_*\}$ and $\mathcal{B}' = \mathcal{B} \cup \{b_*\}$, respectively.

Let us define $T = \mathtt{max}(T', T'')$, i.e., the maximum length of any input or output trace in the unfolded tree of the automaton. The function $\mathtt{Labels}(\mathcal{I}, t) : \text{IIHS} \times \{1, \ldots, T\} \to \wp(\mathcal{L})$ from an IIHS $\mathcal{I}$ and a given level $1 \leq t \leq T$ of its unfolded tree to the set $\mathcal{L}$ of input symbols, output symbols and unobservable symbols of $\mathcal{I}$. Informally, $\mathtt{Labels}(\mathcal{I}, t)$ is the set of all labels of transitions that can be performed with a non-zero probablity from the states at the t^{th} level of the automaton of $\mathcal{I}$.

Definition 14. *For an* IIHS $\mathcal{I} = (M, \mathcal{A}, \mathcal{B}, \mathcal{L}_\tau)$ *with* $M = (Q, \mathcal{L}, \hat{s}, \vartheta)$*, where* $\mathcal{L} = \mathcal{A} \cup \mathcal{B} \cup \mathcal{L}_\tau$*, and for* $t \geq 0$*:*

$$\mathtt{Labels}(\mathcal{I}, t) \equiv \{\ell \in \mathcal{L} \mid \exists \sigma, s \,.\, |\sigma| = t,\ \mathrm{last}(\sigma) \xrightarrow{\ell} s\}$$

The process of normalization of a tree relies on the fact that it is possible to construct an equivalent IIHS $\mathcal{I}' = (M', \mathcal{A}', \mathcal{B}', \mathcal{L}_\tau)$, where $M' = (Q', \mathcal{L}', \hat{s}', \vartheta')$ such that $\mathcal{L}' = \mathcal{A}' \cup \mathcal{B}' \cup \mathcal{L}_\tau$ and its unfolded tree up to depth $2T$ respects, for every $1 \leq t \leq T$:

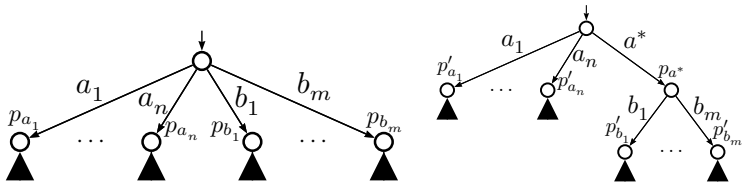
1. $\mathtt{Labels}(\mathcal{I}', t) \cap \mathcal{A}' = \emptyset$ or $\mathtt{Labels}(\mathcal{I}', t) \cap \mathcal{B}' = \emptyset$;
2. $\mathcal{A}' \subseteq \mathtt{Labels}(\mathcal{I}', t)$ or $\mathcal{B}' \subseteq \mathtt{Labels}(\mathcal{I}', t)$;
3. $\mathcal{A}' \subseteq \mathtt{Labels}(\mathcal{I}', t)$ iff $\mathcal{B}' \subseteq \mathtt{Labels}(\mathcal{I}', t+1)$, where we consider the arithmetic on t modulo $2T$;
4. $\mathcal{A}' \subseteq \mathtt{Labels}(\mathcal{I}', 1)$;
5. $|\, trace_{\mathcal{A}'}(\sigma)| = |\, trace_{\mathcal{B}'}(\sigma)| = T$, for all path σ in the unfolded tree.

Condition 1 states that each level can admit input actions or output actions, but not both. Condition 2 states that all input actions need to be listed in an input level, and the same for output levels and actions (as we will see soon, even if we need to associate probability zero to an action). Condition 3 states that input and output levels must necessarily alternate. Condition 4 assures that we always start with an input level. Condition 5 assures that all the leaves of the unfolded tree are in the same level, i.e., the tree is *balanced*.

The proof is straightforward, but we shall give an intuition of it. First, the new symbols a_* and b_* are place holders for the absence of an input and output symbol, respectively. Now, if in a given level t we want to have only input symbols, we can postpone output symbols by adding a_* to the level and "moving" all the output symbols to the subtree of a_*. Figure 6 exemplifies the local transformations we desire in a tree.

Note that in 6(b) the introduction of new nodes changed the probabilities. In general, if we are in an input level, we need to introduce a_* to postpone the output symbols, and the probabilities change as follows:

1. For every a_i, $1 \leq i \leq n$, the associated is probability is mantained as $p'_{a_i} = p_{a_i}$;
2. The probability of the new symbol a_* is introduced as $p_{a_*} = \sum_{i=0}^{m} p_{b_i}$;
3. If $p_{a_*} \neq 0$, then for $1 \leq i \leq m$, the associated probability of b_j is updated to $p'_{b_j} = p_{b_j}/p_{a_*} = p_{b_j}/\sum_{k=0}^{m} p_{b_k}$. If $p_{a_*} = 0$, then $p'_{b_j} = 0$, for $1 \leq i \leq m$, and $p_{b_*} = 1$.



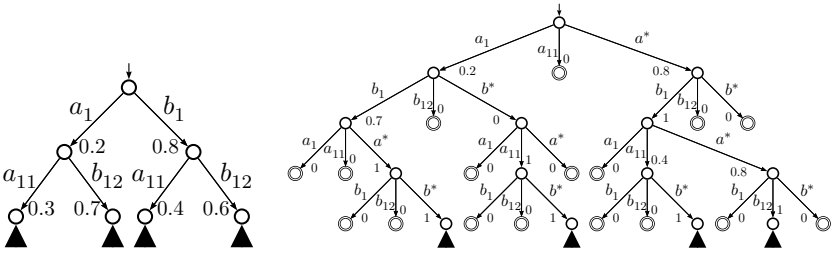
(a) Local nodes of the tree before the transformation

(b) Local nodes of the tree after the transformation

Fig. 6. Local transformation on an IIHS tree

The subtrees of each node of the original tree are preserved as they are, until we apply the same transformation to them. If a node does not have a sutree (i.e, no descendants), we create a subtree by adding all the possible actions in $\mathcal{B}$ with probability 0, and the action b_* with probability 1.

If we are in an output level, the same rules apply, guarding the proper symmetry between input and outputs. We proceed with the same transformation on the next levels of the tree. Figure 7 shows an example of a full transformation on a tree (for the sake of readability, we omit the levels where only $a_* = 1$ or $b_* = 1$).



(a) Tree before transformation

(b) Tree after transformation

Fig. 7. Transformation on an IIHS tree