Quantitative Information Flow in Interactive Systems*

Mário S. Alvim[1] Miguel E. Andrés[2,†] 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 We show that the information-theoretic approach which interprets such systems as (simple) noisy channels is no longer valid. However, the principle can be recovered if we consider channels of a more complicated kind, 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 channels. Furthermore, we show that the capacity of the channels associated to such systems is a continuous function with respect to a pseudometric based on the Kantorovich metric.

keywords: Probabilistic systems, Information flow and Anonymity, Information-theoretic channel, Kantorovich metric.

1 Introduction

Information leakage refers to the problem that arises when the observable behavior of a system reveals information that we would like to keep secret. This is also known as the problem of information flow from *high* variables to *low* variables. In recent years there has been a growing interest in quantitative approaches to this problem, because it is often desirable to quantify the partial knowledge of the secrets in terms of a probability distribution. Another reason is that the mechanisms to protect the information may use randomization to obfuscate the relation between the secrets and the observables.

Among the quantitative approaches, some of the most popular ones are based on Information Theory [5, 16, 4, 24, 6]. The idea is to interpret the system as an information-theoretic *channel*, where the secrets are the input and the observables are the output.

*This work has been partially supported by the project ANR-09-BLAN-0345 CPP, by the INRIA DRI Equipe Associée PRINTEMPS, and by the European Union Seventh Framework Programme under grant agreement no. 295261 (MEALS).

†Supported by the NWO project 612.000.526. Part of the work of Miguel E. Andrés was done at the École Polytechnique, first as a visitor supported by INRIA, and then as a postdoc, supported by a LIX/Qualcomm fellowship.

‡Corresponding author: Catuscia Palamidessi, LIX, École Polytechnique, Rue de Saclay, 91128 Palaiseau Cedex, FRANCE, +33 (0)1 69 33 41 17, catuscia@lix.polytechnique.fr.

The channel matrix consists of the conditional probabilities $p(b \mid a)$, defined as the measure of the executions producing the observable b, relative to 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 works cited above, the secret value is assumed to be chosen at the beginning of the computation. We are interested in the more general scenario in which secrets can be chosen at any point. More precisely, we consider *interactive systems*, i.e. systems in which the generation of secrets and the occurrence of observables can alternate during the computation and influence each other. Examples of interactive systems include *auction protocols* like [32, 27, 25]. Some of these have become very popular thanks to their integration in Internet-based electronic commerce platforms [10, 11, 19]. Other examples of interactive programs include web servers, GUI applications, and command-line programs [3].

In this paper we investigate the applicability of the information-theoretic approach to interactive systems. In order to derive an information-theoretic channel, at a first glance it would seem natural to define the channel matrix by using the definition of $p(b \mid a)$ in terms of the joint and marginal probabilities $p(a, b)$ and $p(b)$. Namely, the entry $p(b \mid a)$ would be defined as the measure of the traces with (secret, observable)-projection (a, b), divided by the measure of the traces with secret projection a. An approach of this kind was proposed in [9]. However, in the interactive case this construction does not really produce an information-theoretic channel. In fact, by definition a channel should be invariant with respect to the input distribution, and this is not the case here, 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 it, with a certain probability. For simplicity we assume that the buyers' offers are mutually exclusive. We assume that the offers are observables, in the sense that they are made public on the website, while the identity of the buyer that actually buys the product should be kept secret from an external observer. The symbols r, q_1, q_2, $\overline{r}$, $\overline{q_1}$, $\overline{q_2}$ represent probabilities, with the convention that $\overline{r} = 1 - r$ (and the same for the pairs q_1, $\overline{q_1}$ and q_2, $\overline{q_2}$). Following [9] we can compute the conditional probabilities as $p(b|a) = \frac{p(a,b)}{p(a)}$, thus obtaining the matrix in Table 1. The matrix however is not invariant with respect to the input distribution. For instance if $r = \overline{r} = \frac{1}{2}$, $q_1 = \frac{2}{3}$, and $q_2 = \frac{1}{3}$ we obtain the matrix in Table 2(a).

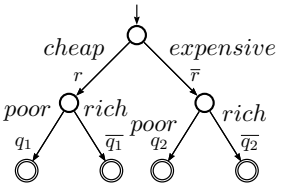


Figure 1: Interactive syst.

If we change the input distribution, for instance by changing the value of q_2 to be $\frac{1}{6}$, also the matrix changes. We obtain, indeed, the new matrix illustrated in Table 2(b).

	cheap	*expensive*
poor	$\frac{rq_1}{rq_1+\overline{r}q_2}$	$\frac{\overline{r}q_2}{rq_1+\overline{r}q_2}$
rich	$\frac{r\overline{q_1}}{r\overline{q_1}+\overline{r}\overline{q_2}}$	$\frac{\overline{r}\overline{q_2}}{r\overline{q_1}+\overline{r}\overline{q_2}}$

Table 1: Matrix of conditional probabilities

Consequently, when the secrets occur *after* the observables and *depend on them*, we cannot consider the conditional probabilities (of the observables given the

(a) $r = \frac{1}{2}, q_1 = \frac{2}{3}, \rho = \frac{1}{2}, q_2 = \frac{1}{3}$

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

(b) $r = \frac{1}{2}, q_1 = \frac{2}{3}, \rho = \frac{1}{4}, q_2 = \frac{1}{6}$

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

Table 2: The channel matrices induced by two input distributions for Example 1

secrets) as representing a classical channel from secrets to observables, and we cannot apply the standard information-theoretic concepts. In particular, we cannot use "the capacity of the matrix" (defined by considering the matrix as a channel matrix, and taking the maximum mutual information over all possible inputs) because in general the maximum is given by a distribution different from the one that has originated the matrix, hence the result would be unsound.

The first contribution of this paper is to consider an extension of the theory of channels which makes the information-theoretic approach applicable also in 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. Recent results in Information Theory [30] have shown that, in such channels, the transmission rate does not correspond to the maximum mutual information (the standard notion of capacity), but rather to the maximum normalized *directed information*, a concept introduced by Massey [17]. We propose to adopt this latter notion to represent leakage.

Our model of attacker is the interactive version of the attacker associated to Shannon entropy in the classification of Köpf and Basin [15], based on an eavesdropper scenario. We recall that in [15] an attacker is defined by the kind of questions that he can pose to a hypothetical oracle. In the case of Shannon entropy the questions are of the form "does s belong to S?" where s is the secret that the attacker is trying to figure out, and S is a subset of the domain of secret values. The degree of invulnerability of the secret is the average number of questions that the attacker needs to ask in order to find out the exact value of the secret, under the best strategy (i.e. the best choice of the S's) for the given probability distribution on the secret values. It is easy to see that the invulnerability degree corresponds to the Shannon entropy of the secret. In the case of a standard single-use channel, the invulnerability degree of the secret *before* the attacker observes the output is the entropy of the input, determined by its a priori distribution. The invulnerability degree *after* the attacker observes the output is the conditional entropy of the input given the output, determined by its a posteriori distri-

bution. The latter is always smaller than or equal to the first. The difference between these invulnerability degrees corresponds to the mutual information, and represents the leakage of the system.

In our interactive framework we consider the same scenario, but iterated. At each time step, we consider the input sequence so far; and the increase of its vulnerability caused by the observation of the new output is given by the contribution of the present step to the leakage. The sum of all these contributions represents the total leakage and, as we will see, corresponds to Massey's directed information. We will come back to the model of attacker in Section 5, and discuss also a variant of this interpretation.

Gray investigated a concept similar to directed information in [13]. In contrast to our model, which is based on an eavesdropper scenario, he considered leakage in a sender-receiver model. More precisely, he considered a system based on Millen's synchronous state machine [20], and connected to "low" and "high" environments via communication channels. His purpose was to measure the flow of information from the high environment to the low one, assuming that the only way for the low environment to learn about the high one (and vice versa) is through the system. To this end, he defined a notion of "quasi-directed information" by extending Gallager's formula for discrete finite state channels [12]. He also conjectured a correspondence between the quasi-directed information and the transmission rate of the channel. His formulation of quasi-directed information, however, is not completely the same as directed information, and as a result the conjecture does not hold. We come back to this point in Section 5, after Definition 9.

A second contribution of our work is the proof that the channel capacity is a continuous function of a pseudometric on interactive systems based on the Kantorovich metric. The reason why we are interested in the continuity of the capacity is for computability purposes. Given a function f from a (pseudo)metric space X to a (pseudo)metric space Y the continuity of f means that, given a sequence of objects $x_1, x_2, \ldots \in \mathcal{X}$ converging to $x \in \mathcal{X}$, the sequence $f(x_1), f(x_2), \ldots \in \mathcal{Y}$ converges to $f(x) \in \mathcal{Y}$. Hence $f(x)$ can be approximated by the objects $f(x_1), f(x_2), \ldots$. The typical use of this property is in the case of execution trees generated by programs containing loops. Generally the automaton expressing the semantics of the program can be seen as the (metric) limit of the sequence of trees generated by unfolding the loop at an increasingly deeper level. The continuity of the capacity means that we can approximate the real capacity by the capacities of these trees.

The continuity of the channel capacity was also proved in [9] for simple channels, but the proof does not adapt to the case of channels with memory and feedback and we had to devise a different technique. We illustrate this point by showing a counterexample (cfr. Example 6).

1.1 Plan of the paper

The paper is organized as follows. Section 2 reviews some important concepts from Probabilistic Automata and Information Theory. Section 3 reviews the notion of channel with memory and feedback that is the core of the model we propose. We discuss the concept of directed information and also the concept of capacity in the presence of feedback. Section 4 contains our main contribution. We explain how Interactive

Information Hiding Systems (IIHSs) can be modeled using channels with memory and feedback. In particular we show that for any IIHS there is always a channel that simulates its probabilistic behavior. In Section 5 we discuss our notion of adversary and we define the quantification of information leakage as the channel's directed information from input to output, or as the directed capacity, depending on whether the input distribution is fixed or not. In Section 6 we show an example of our model applied to a protocol: the Cocaine Auction protocol. Section 7 proposes a pseudometric structure on IIHSs based on the Kantorovich metric. We also show that the capacity of the channels associated to interactive systems is a continuous function with respect to this pseudometric. In Sections 8 and 9 we review and discuss the main results of the paper and illustrate some future work.

A preliminary version of this paper appeared in the proceedings of CONCUR 2010 [1]. The additional material presented here consists in the proofs, the auxiliary Lemmata 4, 7, and 9, Propositions 8 and 10, more examples, and a more elaborate discussion about the model.

2 Preliminaries

In this section we briefly review some basic notions that we will need throughout 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 $\sum_{s\in\mathcal{S}} \mu(s) = 1$. The set of all discrete probability distributions on $\mathcal{S}$ is $\mathcal{D}(\mathcal{S})$.

A *probabilistic automaton* [22] 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. In such automata, when $\vartheta(s) \neq \emptyset$, we overload the notation and denote by $\vartheta(s)$ the distribution outgoing from s.

A *path* in a probabilistic automaton is a sequence $\sigma = s_0 \xrightarrow{\ell_1} s_1 \xrightarrow{\ell_2} \cdots$ where $s_i \in \mathcal{S}$, $\ell_i \in \mathcal{L}$ and $s_i \xrightarrow{\ell_{i+1}} s_{i+1}$. A path can be *finite* in which case it ends with a state. A path is *complete* if it is either infinite, or finite ending in a terminal state. Given a finite path σ, $\mathrm{last}(\sigma)$ denotes its last state. Let $\mathrm{Paths}_s(M)$ denote the set of all paths, $\mathrm{Paths}^\star{}_s(M)$ the set of all finite paths, and $\mathrm{CPaths}_s(M)$ the set of all complete paths of an automaton M, starting from the state s. We will omit s if $s = \hat{s}$. Paths are ordered by the prefix relation, which we denote by $\leq$. The *trace* of a path is the sequence of actions in $\mathcal{L}^* \cup \mathcal{L}^\infty$ obtained by removing the states, hence for the above σ we have $\mathit{trace}(\sigma) = l_1 l_2 \ldots$. If $\mathcal{L}' \subseteq \mathcal{L}$, then $\mathit{trace}_{\mathcal{L}'}(\sigma)$ is the projection of $\mathit{trace}(\sigma)$ on the elements of $\mathcal{L}'$.

Let $M = (\mathcal{S}, \mathcal{L}, \hat{s}, \vartheta)$ be a (fully) probabilistic automaton, $s \in \mathcal{S}$ a state, and let $\sigma \in \mathrm{Paths}^\star_s(M)$ be a finite path starting in s. The *cone* generated by σ is the set of

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

Let $\Omega_s \triangleq \mathrm{CPaths}_s(M)$ be the sample space, and let $\mathcal{F}_s$ be the smallest σ-algebra induced by the cones generated by all the finite paths of M. 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 [7]. Let A, B denote two discrete random variables with finitely many values, and 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 $H(A) = -\sum_{\mathcal{A}} p(a) \log p(a)$ and it measures the uncertainty of A. It takes its minimum value $H(A) = 0$ when $p_A(\cdot)$ is a point mass (also called Dirac measure). 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 $H(A|B) = -\sum_{\mathcal{B}} p(b) \sum_{\mathcal{A}} p(a|b) \log p(a|b)$, and it measures the uncertainty of A when B is known. It is well-known 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 $0 \leq I(A; B) \leq H(A)$. If C is a third random variable, the *conditional mutual information* between A and B given C is defined as $I(A; B|C) = H(A|C) - H(A|B, C)$.

The (conditional) entropy and mutual information respect the *chain rules*. Namely, given the random variables $A_1, A_2, \ldots, A_k$, B and C, we have:

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

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

A (*discrete memoryless*) *channel* is a tuple $(\mathcal{A}, \mathcal{B}, p(\cdot|\cdot))$, where $\mathcal{A}, \mathcal{B}$ are the sets of input and output symbols, respectively, and $p(b|a)$ is the probability of observing the output symbol b when the input symbol is a. These conditional probabilities constitute the *channel matrix*. An input distribution $p_A(\cdot)$ over $\mathcal{A}$ together with the channel determine the joint distribution $p(a, b) = p(a|b) \cdot p(a)$ and consequently $I(A; B)$.

The maximum $I(A;B)$ over all possible input distributions is the channel's *capacity*. Finally, a family $\rho = \{p_v(\cdot)\}_v$ of probability measures parametrized on v is called a *stochastic kernel*[1].

3 Discrete channels with memory and feedback

In this section we present the notion of channel with memory and feedback. We assume a scenario in which the channel is used repeatedly, in a finite temporal sequence of steps $1, \ldots, T$. Intuitively, memory means that the output at time t depends on the input and output histories, i.e. on the inputs up to time t, and on the output up to time $t-1$. Feedback means that the input at time t depends on the outputs up to time $t-1$.

We adopt the following notation, which appears to be standard in the literature of channels with memory and feedback.

Convention 1. Given a set of symbols (alphabet) $\mathcal{A} = \{a_1, \ldots, a_n\}$, we use a Greek letter (α, β, ...) to denote a sequence of symbols ordered in time. Given a sequence $\alpha = a_{i_1} a_{i_2} \ldots a_{i_m}$, the notation α_t represents the symbol at time t, i.e. a_{i_t}, while α^t represents 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 $\alpha^2 = a_3 a_7$. Analogously, if X is a random variable, then X^t denotes the sequence of t consecutive instances $X_1, \ldots, X_t$ of X.

We now define formally the concepts of memory and feedback. Consider a channel from input A to output B. The channel behavior after T uses can be fully described by the joint distribution of $A^T \times B^T$, namely by the probabilities $p(\alpha^T, \beta^T)$. Using the chain rule, we can decompose these probabilities as follows:

$$p(\alpha^T, \beta^T) \quad = \quad \prod_{t=1}^{T} p(\alpha_t | \alpha^{t-1}, \beta^{t-1}) p(\beta_t | \alpha^t, \beta^{t-1}) \tag{3}$$

Definition 1. We say that the channel *has feedback* if, in general, $p(\alpha_t | \alpha^{t-1}, \beta^{t-1}) \neq p(\alpha_t | \alpha^{t-1})$, i.e. the probability of α_t depends not only on α^{t-1}, but also on β^{t-1}. Analogously, we say that the channel *has memory* if, in general, $p(\beta_t | \alpha^t, \beta^{t-1}) \neq p(\beta_t | \alpha_t)$, i.e. the probability of β_t depends on α^t and β^{t-1}.

Note that in the opposite case, i.e. when $p(\alpha_t | \alpha^{t-1}, \beta^{t-1})$ coincides with $p(\alpha_t | \alpha^{t-1})$ and $p(\beta_t | \alpha^t, \beta^{t-1})$ coincides with $p(\beta_t | \alpha_t)$, then we have a classical channel (memoryless, and without feedback), in which each use is independent from the previous ones. The only possible dependency on the history is the one of a_t on a^{t-1}. This is because $A_1, \ldots, A_T$ are in general correlated, due to the fact that they are produced by an encoding function. Note that in absence of memory and feedback (3) reduces to $p(\alpha^T, \beta^T) = \prod_{t=1}^{T} p(\alpha_t | \alpha^{t-1})\, p(\beta_t | \alpha_t)$, which is the standard formula for a classical channel after T uses.

The above is a very abstract description of a channel with memory and feedback. We now discuss a more concrete notion following the presentation of [30] . Such

[1]The general definition of stochastic kernel is more complicated (cfr. [30]), but it reduces to this one in the case of discrete channels, which is what we use in this paper.

a channel, represented in Figure 2, consists of a sequence of components formally defined as a family of stochastic kernels $\{p(\cdot\,|\alpha^t, \beta^{t-1})\}_{t=1}^{T}$ over $\mathcal{B}$. The probabilities $p(\beta_t|\alpha^t, \beta^{t-1})$ represent the *innermost behavior* of the channel at time t, $1 \leq t \leq T$: the internal channel takes the input α_t and, depending on the history of inputs and outputs so far, it produces an output symbol β_t. The output is then fed back to the encoder with delay one. On the output 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 according to a code function φ_t (we will explain this concept in the next paragraph). 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:

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

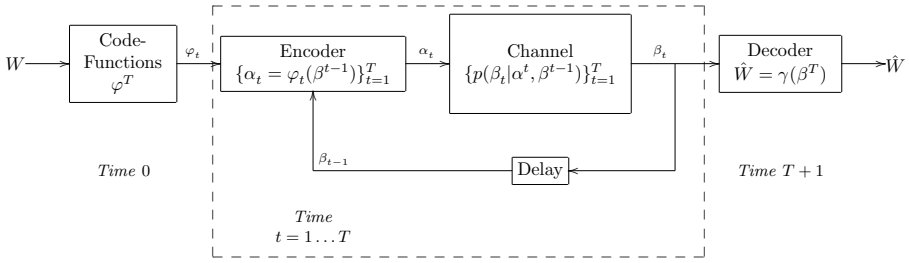


Figure 2: Model for discrete channel with memory and feedback

Let us now explain the concept of code function. Intuitively, a code function is a strategy to encode the message into a suitable representation to be transmitted through the channel. There is a code function for each possible message, and the function is fixed at the very beginning of the transmission (time $t = 0$). However, the encoding can use the information provided via feedback, so each component φ_t ($1 \leq t \leq T$) of the code function takes as parameter the history of feedback β^{t-1} to generate the next input symbol α_t.

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

3.1 The power of feedback

The original purpose of *communication channel* models is to represent data transmission from a source to a receiver. Shannon's Channel Coding Theorem states for every channel there is an encoding scheme that allows a transmission rate arbitrarily close to the channel capacity with a negligible probability of error (if the number of uses of the channel is large enough). A general way to find an optimal encoding scheme that is also easy to decode has not been found yet. The use of feedback, however, can simplify the design of both the encoder and the decoder. The following example illustrates the idea.

Example 2. Consider a discrete memoryless binary channel $\{\mathcal{A}, \mathcal{B}, p(.|.)\}$ with $\mathcal{A} = \{0, 1\}$, $\mathcal{B} = \{0, 1, \mathtt{e}\}$ and the channel matrix of Table 3.

	0	1	e
0	0.8	0	0.2
1	0	0.8	0.2

Table 3: Channel matrix for binary erasure channel

This kind of channel is called *erasure channel* because it can lose (or *erase*) bits during the transmission with a certain probability. Namely, any bit has 0.8 probability of being correctly transmitted, and 0.2 probability of being lost. On the output side the encoder is able to detect whether the bit was erased (by receiving an $\mathtt{e}$ symbol), but it cannot tell which was the actual value of the original bit. The Channel Coding 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.

Following simple principles described in [7], an encoding that achieves the capacity can be easily obtained if the channel can be used with feedback. The idea is an adaptation of the stop-and-wait protocol [26, 28]. Suppose that every bit received on the output end of the channel is fed back noiselessly to the source with delay 1. Define the encoding as follows: for each bit transmitted, the encoder checks via feedback if the bit was erased. If not, the encoder moves on to transmit the text of the message. If yes, the encoder transmits the same bit again.

It is easy to see that with this encoding scheme the transmission rate is 0.8 bit per usage of the channel, since in 80% of the cases the bit is transmitted properly, and in 20% it is lost and a retransmission is needed.

In Appendix A we come back to this example to illustrate in more detail the design and the function of the encoder and decoder. Note that the channel capacity in the above example does not increase with the addition of feedback (it is 0.8 bit per usage of the channel with or without feedback). This is because the channel is memoryless: *feedback does not increase the capacity of discrete memoryless channels* [7]. In general however, feedback *does* increase the capacity.

3.2 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 Channel Coding Theorem, can be obtained as the supremum of the mutual information over all possible input distributions. In the presence

of feedback, however, this correspondence no longer holds. More specifically, mutual information no longer represents 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 (see Example 5). However, feedback, i.e the way the output influences the next input, is not 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* [17].

Definition 2. 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 5 we shall discuss the 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 in presence of feedback. For finite T, the capacity of a channel with memory and feedback is:

$$C_T = \sup_{\mathcal{D}_T} \frac{1}{T} I(A^T \to B^T) \tag{5}$$

The capacity is also defined when T is infinite, see [30]. However in this paper we consider only the finite case.

4 Interactive systems as channels with memory and feedback

Interactive Information Hiding Systems (IIHS) were introduced in [2] to represent systems where secrets (inputs) and observable (outputs) can interleave and influence each other. They are a variant of probabilistic automata in which actions are divided in secrets and observables. They can be of two kinds: *fully probabilistic*, and *secret-nondeterministic* (or *input-nondeterministic*). In the former there is no nondeterminism, while in the latter every secret choice is fully nondeterministic.

In this paper we consider *normalized* IIHSs, in which secrets and observables alternate, and the actions at the first level are secrets. We note that this is not really a restriction, because given an IIHS which is not normalized, it is always possible to transform it into a normalized IIHS which is equivalent to the former one up to a given execution level. The reader can find in Appendix B the formal definition of the transformation. Furthermore, we require that for each state s and each action ℓ there is at most one state that can be reached from s by performing an ℓ transition.

We give now the formal definition of the kind of IIHSs that we will use in this paper.

Definition 3. A (normalized) IIHS is a triple $\mathcal{I} = (M, \mathcal{A}, \mathcal{B})$, where $\mathcal{A}$ and $\mathcal{B}$ are disjoint sets of secrets and observables respectively, M is a probabilistic automaton $(\mathcal{S}, \mathcal{L}, \hat{s}, \vartheta)$ with $\mathcal{L} = \mathcal{A} \cup \mathcal{B}$, and, for each $s \in \mathcal{S}$:

1. either $\vartheta(s) \subseteq \mathcal{D}(\mathcal{A} \times \mathcal{S})$ or $\vartheta(s) \subseteq \mathcal{D}(\mathcal{B} \times \mathcal{S})$. We call s *a secret state* in the first case, and *an observable state* in the second case;

2. if $s \xrightarrow{\ell} r$ then: if s is a secret state then r is an observable state, and if s is an observable state then r is a secret state;

3. $\hat{s}$ is a secret state;

4. if s is an observable state then $|\vartheta(s)| \leq 1$;

5. either:

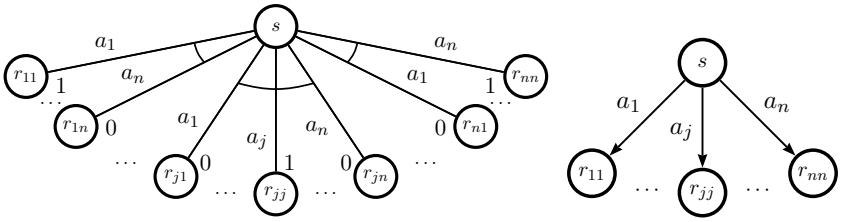
   (i) for every secret state s we have $|\vartheta(s)| \leq 1$ *(fully probabilistic* IIHS*)*,

   or

   (ii) for every secret state s there exist a_i and s_i $(i = 1, \ldots, n)$ such that $\vartheta(s) = \{\delta(a_i, s_i)\}_{i=1}^{n}$, where $\delta(a_i, s_i)$ is the Dirac measure *(secret-nondeterministic* IIHS*);*

6. for every state s and action ℓ there exists a unique state r such that $s \xrightarrow{\ell} r$.

In the rest of the paper we will omit the adjective "normalized" for simplicity. In the above definition, Conditions 1 and 2 imply that the IIHS is alternating between secrets and observables. Moreover, all the transitions between nodes at two consecutive depths have either secret actions only, or observable actions only. Condition 3 means that the first level contains secret actions. Condition 4 means that all observable transitions are fully probabilistic. Condition 5 means that all secret transitions are either fully probabilistic or fully nondeterministic. The term "nondeterministic" is justified by the fact that the scheme of Condition 5(ii), represented in Figure 3(a), is equivalent to the one of Figure 3(b).



(a) Nondeterministic input using Dirac measures (b) Equivalent scheme

Figure 3: Scheme of secret transitions for secret-nondeterministic IIHSs.

Note that we do not consider here internal nondeterminism such as that one arising from interleaving of concurrent processes. This means that we make a rather restricted use of probabilistic automata, but this is enough for our purposes. The nondeterminism generated by concurrency gives rise to a new set of problems (see for example [4]) which are orthogonal to those considered in this paper.

Condition 6 means that the secret and observable actions determine the states. As a consequence, the actions are enough to retrieve the path. This is expressed by the following proposition:

Proposition 2. *Given an* IIHS*, consider two paths σ and σ'. If $trace_{\mathcal{A}}(\sigma) = trace_{\mathcal{A}}(\sigma')$ and $trace_{\mathcal{B}}(\sigma) = trace_{\mathcal{B}}(\sigma')$, then $\sigma = \sigma'$.*

Proof. By induction on the length of the traces. The initial state of the automaton is uniquely determined by the empty (secret and observable) traces. Assume now we are in a state s uniquely determined by secret and observable traces α and β, respectively. If s makes a secret transition $s \xrightarrow{a} s'$, then by Condition 6 there is only one state s' reachable from s via an a-transition, and therefore s' is uniquely determined by the secret trace $\alpha' = \alpha a$ and the observable trace β. The case in which s makes an observable transition is similar. □

4.1 Construction of the channel associated to an IIHS

We now show how to associate a channel to an IIHS.

In an interactive system secrets and observables may interleave and influence each other. Considering a channel with memory and feedback is a way to capture this rich 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 feedback. This alternating mutual influence between secrets and observables can be modeled by repeated uses of the channel. Each time the channel is used it represents a different state of the computation, and the conditional probabilities of observables with respect to secrets can depend on this state. The addition of memory to the model allows expressing the dependency of the channel matrix on such a state.

We will see that a secret-nondeterministic IIHS determines a channel as specified by its stochastic kernel, while a fully probabilistic IIHS determines, additionally, the input distribution. In Section 6 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. Let $\mathcal{I}$ be an IIHS. For each t, the channel's stochastic kernel corresponding to $\mathcal{I}$ is defined as $p(\beta_t|\alpha^t, \beta^{t-1}) = \vartheta(s)(\beta_t, s')$, where s is the state reached from the root via the path σ whose secret and observable traces are α^t and β^{t-1} respectively.

Note that s and s' in the previous definition are well defined: by Proposition 2, s is unique, and since the choice of β_t is fully probabilistic, s' is also unique.

The following example illustrates how to apply Definition 4, with the help of Proposition 2, to build the channel matrix of a simple example.

Example 3. 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 observable, 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 that, after normalization, $T = 3$. Figure 4 shows an automaton for this example in normalized form. Transitions with null probability are omitted, and the symbol a_* is used as a place holder to achieve the normalized IIHS (see Appendix).

To construct the stochastic kernels $\{p(\beta_t|\alpha^t, \beta^{t-1})\}_{t=1}^T$, 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 the 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 $p(\beta_2 = cheap|\alpha^2 = a_*, poor, \beta^1 = expensive) = \vartheta(s)(cheap, s')$. By Proposition 2, the traces $\alpha^2 = a_*, poor, \beta^1 = expensive$ determine a unique state s in the automaton, namely, the state $s = 5$. Moreover, from the state 5 a unique transition labelled with the action $cheap$ is possible, leading to the state $s' = 11$. Therefore, we can conclude that $p(\beta_2 = cheap|\alpha^2 = a_*, poor, \beta^1 = expensive) = \vartheta(s = 5)(cheap, s' = 11) = p_{23}$.

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

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

Definition 5. Let $\mathcal{I}$ be an IIHS. If $\mathcal{I}$ is fully probabilistic, the associated channel has a conditional input distribution for each t defined as $p(\alpha_t|\alpha^{t-1}, \beta^{t-1}) = \vartheta(s)(\alpha_t, s')$, where s is the state reached from the root via the path σ whose secret and observable traces are α^{t-1} and β^{t-1} respectively.

Example 4. Since the system of Example 3 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, for instance, 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 5, we see that $p(\alpha_2 = poor|\alpha_1 = a_*, \beta^1 = expensive) = \vartheta(s)(poor, s')$. By Proposition 2, the traces $\alpha^1 = a_*, \beta^1 = expensive$ determine a unique state s in the automaton, namely, the state $s = 2$. Moreover, from the state 2 a unique transition labelled with the action $poor$ is possible, leading to the state $s' = 5$. Therefore, we can conclude that $p(\alpha_2 = poor|\alpha_1 = a_*, \beta^1 = expensive) = \vartheta(s = 2)(poor, s' = 5) = q_{12}$.

Similarly, with $t = 3$ and history $\alpha^2 = a_*, rich, \beta^2 = cheap, expensive$, the secret symbol $\alpha_3 = rich$ can be observed with probability $p(\alpha_3 = rich|\alpha^2 = \alpha_*, rich, \beta^0 = cheap, expensive) = \vartheta(s = 10)(cheap, s' = 22) = \overline{q_{24}}$.

4.2 Lifting the channel inputs to reaction functions

Definitions 4 and 5 show how to obtain the the joint probabilities $p(\alpha^t, \beta^t)$ for a fully probabilistic IIHS. We still need to show in what sense this joint probability distribution defines an information-theoretic channel.

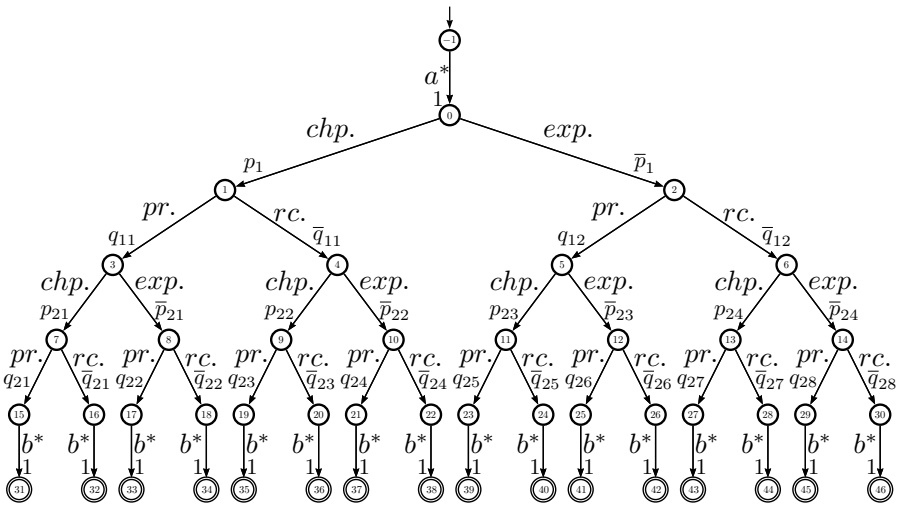


Figure 4: The normalized IIHS for the extended website example

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 probabilities $\{p(\alpha_t|\alpha^{t-1}, \beta^{t-1})\}_{t=1}^T$. In an information-theoretic channel, the value of α_t is determined in the encoder by a deterministic function $\varphi_t(\beta^{t-1})$. Therefore, 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.

As shown in [30], 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 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 3 ([30]). *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:*

$$Q(\beta_t|\varphi^t,\beta^{t-1}) = p(\beta_t|\varphi^t(\beta^{t-1}),\beta^{t-1}) \tag{6}$$

Since in our setting the concept of encoder makes little sense as there is no information to encode, we externalize the probabilistic behavior of α_t as follows. Code functions become *a single set of reaction functions $\{\varphi_t\}_{t=1}^T$ with β^{t-1} as parameter* (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 sequence of stochastic kernels $\{p(\varphi_t|\varphi^{t-1})\}_{t=1}^T$. 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 4. *Let $\mathcal{X},\mathcal{Y}$ be non-empty 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: $\mathcal{X}=\{\tilde{x}\}$. 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 $\tilde{x}\in\mathcal{X}'$ and $\mathring{x}\notin\mathcal{X}'$. 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. *Let* $\mathfrak{I}$ *be a fully probabilistic* IIHS *inducing the joint probability distribution* $p(\alpha^t,\beta^t)$*,* $1 \le t \le T$*, on secret and observable traces. It is always possible to construct a channel with memory and feedback, and an associated probability distribution* $Q(\varphi^T,\alpha^T,\beta^T)$*, which corresponds to* $\mathfrak{I}$ *in the sense that, for every* $1 \le t \le T$*,* α^t*,* β^t*, the equality* $Q(\alpha^t,\beta^t) = p(\alpha^t,\beta^t)$ *holds.*

Proof. First of all we note that, by the laws of probability, $Q(\alpha^t,\beta^t) = \sum_{\varphi^t} Q(\varphi^t,\alpha^t,\beta^t)$. 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:

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

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 4.

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 4)

$$\sum_{\varphi^{t-1}} Q(\varphi^{t-1}, \alpha^{t-1}, \beta^{t-1}) \cdot p(\beta_t|\alpha^t, \beta^{t-1}) \cdot p(\alpha_t|\alpha^{t-1}, \beta^{t-1})$$

$=$

$$p(\beta_t|\alpha^t, \beta^{t-1}) \cdot p(\alpha_t|\alpha^{t-1}, \beta^{t-1}) \cdot \sum_{\varphi^{t-1}} Q(\varphi^{t-1}, \alpha^{t-1}, \beta^{t-1})$$

$=$ (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 6. *Let $\mathfrak{I}$ be a fully probabilistic* IIHS. *Let $\{p(\beta_t|\alpha^t, \beta^{t-1})\}_{t=1}^T$ be a sequence of stochastic kernels and $\{p(\alpha_t|\alpha^{t-1}, \beta^{t-1})\}_{t=1}^T$ a sequence of input distribu-*

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

$$p(\varphi_1) = p(\alpha_1|\alpha^0, \beta^0) = p(\alpha_1) \quad (7)$$

$$p(\varphi_t|\varphi^{t-1}) = \prod_{\beta^{t-1}} p(\varphi_t(\beta^{t-1})|\varphi^{t-1}(\beta^{t-2}), \beta^{t-1}), \quad 2 \leq t \leq T \quad (8)$$

Figure 5 depicts the model for IIHS. Note that, in relation to Figure 2, there are some simplifications: (1) no message w is needed; (2) the encoder becomes an "interactor"; (3) 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 interactor produces the next input symbol α_t by applying the reaction function φ_t to the fed back output 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, which uses it for producing the next input.

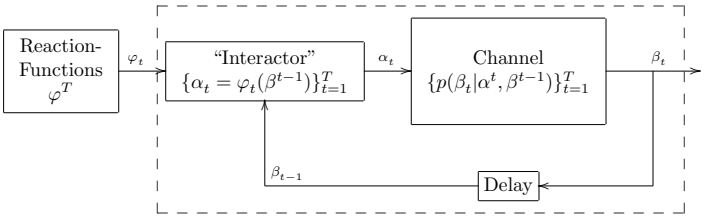


Figure 5: Channel with memory and feedback model for IIHS

We conclude this section by remarking on an intriguing coincidence: The notion of reaction function sequence φ^T, on the IIHSs, corresponds to the notion of deterministic scheduler [22]. In fact, each reaction function φ_t selects the next step, α_t, on the basis of the β^{t-1} and α^{t-1} (generated by φ^{t-1}), and β^{t-1}, α^{t-1} represent the path until that state.

5 Leakage in Interactive Systems

In this section we propose a definition for the notion of leakage in interactive systems. We first argue that mutual information is not the correct notion, and we propose to replace it with the directed information instead.

In the case of channels with memory and feedback, mutual information is defined as $I(A^T; B^T) = H(A^T) - H(A^T|B^T)$, and it is still symmetric (i.e. $I(A^T; B^T) = I(B^T; A^T)$). Since the roles of A^T and B^T in $I(A^T; B^T)$ are interchangeable, this concept cannot capture *causality*, in the sense that it does not imply that A^T causes B^T, nor conversely. Mutual information expresses *correlation* between the sequences of random variables A^T and B^T.

Mathematically, for T usages of the channel, the mutual information $I(A^T; B^T)$ can be expressed with the help of the chain rule of (2) in the following form.

$$I(A^T; B^T) = \sum_{t=1}^{T} I(A^T; B_t | B^{t-1}) \tag{9}$$

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 B^{t-1}. The equation emphasizes 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 A^T, including $A_{t+1}, A_{t+2}, \ldots, A_T$, has a statistical correlation with B_t. Indeed, in the presence of feedback, B_t may influence $A_{t+1}, A_{t+2}, \ldots, A_T$.

In order to show how the concept of directed information contrasts with the above, let us recall its definition:

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

$$I(B^T \to A^T) = \sum_{t=1}^{T} I(A_t; B^{t-1} | A^{t-1}).$$

These notions capture the concept of *causality*, to which the definition of mutual information is indifferent. The correlation between inputs and outputs $I(A^T; B^T)$ is split into the information $I(A^T \to B^T)$ that flows from input to output through the channel and the information $I(B^T \to A^T)$ that flows from output to the input via feedback. Note that the directed information is not symmetric: the flow from A^T to B^T takes into account the correlation between A^t and B_t, while the flow from B^T to A^T takes into account the correlation between B^{t-1} and A_t.

It was proved in [30] that

$$I(A^T; B^T) = I(A^T \to B^T) + I(B^T \to A^T) \tag{10}$$

i.e, the mutual information is the sum of the directed information flow in both senses. Note that this formulation highlights the symmetry of mutual information from yet another perspective.

Once we split mutual information into directed information in the two opposite directions, it is important to understand the different roles that the information flow in each direction plays. $I(A^T \to B^T)$ 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, $I(B^T \to A^T)$ represents how the environment reacts to the system: given the system outputs, the environment produces new inputs. We argue that the information flow from outputs to inputs 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 of the adversary. As a further justification, observe that this is a natural extension of the classical approach, where the choice of secrets is seen as external to the system, i.e. determined by the environment. The probability distribution on the secrets 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 a natural way to do so is to consider the secret choices, at every stage of the computation, as external. Their probability distributions, which are now in general conditional probability distributions depending on the history of secrets and observables, should therefore be considered as part of the external knowledge, and not counted as leakage.

The following example supports our claim that, in the presence of feedback, mutual information is not a correct notion of leakage.

Example 5. Consider the discrete memoryless channel with secret alphabet $\mathcal{A} = \{a_1, a_2\}$ and observable alphabet $\mathcal{B} = \{b_1, b_2\}$ whose matrix is represented in Table 4.

Suppose that the channel is used with feedback, in such a way that, for all $1 \leq t \leq T$, we have $\alpha_{t+1} = a_1$ 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 $I(B^T \to A^T)$.

	b_1	b_2
a_1	0.5	0.5
a_2	0.5	0.5

Table 4: Channel matrix for Example 5

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, making the necessary adaptations.

As discussed in the introduction, in the non interactive case the definition of leakage as mutual information, for a single use of the channel, is

$$I(A; B) = H(A) - H(A|B)$$

(cfr. for instance [4, 15]). This amounts to viewing the leakage as the difference between the a priori invulnerability and the a posteriori one. As explained in the introduction, these are represented by $H(A)$ and $H(A|B)$ respectively. This corresponds to the model of an attacker based on Shannon entropy discussed by Köpf and Basin in [15].

In the interactive case, we can extend this notion by considering the leakage at every step t as given by

$$I(A^t; B_t|B^{t-1}) = H(A^t|B^{t-1}) - H(A^t|B_t, B^{t-1})$$

The notion of attack is the same modulo the fact that we consider all the input from the beginning up to step t, and the difference in its vulnerability induced by the observation of B_t (the output at step t), taking into account the observation history B^{t-1}. It is then natural to consider as total leakage the summation of the contributions $I(A^t; B_t|B^{t-1})$

for all the steps t. This is exactly the notion of directed information (cfr. Definition 2):

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

Definition 8. The information leakage of a fully probabilistic IIHS is defined as the directed information $I(A^T \to B^T)$ of the associated channel with memory and feedback.

We now show an equivalent formulation of directed information that leads to a new interpretation in terms of an attack model. First we need the following lemma.

Lemma 7. $I(B^T \to A^T) = H(A^T) - \sum_{t=1}^{T} H(A_t | A^{t-1}, B^{t-1})$

Proof.

$$\begin{aligned}
& I(B^T \to A^T) \\
= \; & \sum_{t=1}^{T} I(A_t; B^{t-1} | A^{t-1}) && \text{(by Definition 2)} \\
= \; & \sum_{t=1}^{T} \big(H(A_t | A^{t-1}) \\
& - H(A_t | A^{t-1}, B^{t-1}) \big) && \text{(by definition of mutual info.)} \\
= \; & H(A^T) - \sum_{t=1}^{T} H(A_t | A^{t-1}, B^{t-1}) && \text{(by the chain rule)}
\end{aligned}$$ □

The next proposition points out the announced alternative formulation of directed information from input to output:

Proposition 8. $I(A^T \to B^T) = \sum_{t=1}^{T} H(A_t | A^{t-1}, B^{t-1}) - H(A^T | B^T)$

Proof.

$$\begin{aligned}
& I(A^T \to B^T) \\
= \; & I(A^T; B^T) - I(B^T \to A^T) && \text{(by (10))} \\
= \; & I(A^T; B^T) - H(A^T) \\
& + \sum_{t=1}^{T} H(A_t | A^{t-1}, B^{t-1}) && \text{(by Lemma 7)} \\
= \; & H(A^T) - H(A^T | B^T) - H(A^T) \\
& + \sum_{t=1}^{T} H(A_t | A^{t-1}, B^{t-1}) && \text{(by definition of mutual info.)} \\
= \; & \sum_{t=1}^{T} H(A_t | A^{t-1}, B^{t-1}) - H(A^T | B^T)
\end{aligned}$$ □

We note that the term $\sum_{t=1}^{T} H(A_t|A^{t-1}, B^{t-1})$ can be seen as the entropy H_R of the reactor R, i.e. the entropy of the inputs, taking into account their dependency on the previous outputs. This brings us to an intriguing alternative interpretation of leakage:

Remark 1. The leakage can be seen as the difference between the a priori invulnerability degree of the whole secret A^T, assuming that the attacker knows the distribution of the reactor, and the a posteriori invulnerability degree, after the adversary has observed the whole output B^T.

In Section 6 we give an extensive and detailed example of how to calculate the leakage for an actual security protocol.

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

Definition 9. The *maximum leakage* of a secret-nondeterministic IIHS is defined as the capacity C_T of the associated channel with memory and feedback (cfr. (5)).

A comparison with the definition of Gray (cfr. [13], Definition 5.3) is in order. As explained in the introduction, Gray's model is more complicated than ours, because it assumes that low and high variables are present at both ends of the channel. If we restrict the definition of Gray's capacity C^G to our case, by eliminating the low input and the high output, we obtain the following formula:

$$C_T^G = \sup_{\mathcal{D}_T} \frac{1}{T} \sum_{t=1}^{T} I(A^{t-1}; B_t|B^{t-1}) \tag{11}$$

By comparing (5), which is based on Definition 2, to (11), we can see that the only difference is that (11) considers the correlation between B_t andA^{t-1} instead of A^t. This seems to be intentional (cfr. [13], discussion after Definition 4.1). We are not sure why C^G is defined in this way, our best guess is that the high values must be those of the previous time step in order to encompass the theory of McLean [18]. In any case, Gray's conjecture that C_T^G corresponds to the channel transmission rate does not hold. For instance, it is easy to see that for $T = 1$ we always have $C_T^G = 0$, but there obviously are channels which can transmit a non-zero amount of information even with one single use.

We conclude this section by showing that our approach to the notion of leakage generalizes the classical approach (based on mutual information) to the case of feedback. The idea is that, if a channel does not have feedback, then $I(B^T \to A^T) = 0$ and therefore $I(A^T; B^T) = I(A^T \to B^T)$. In our opinion, the fact that mutual information turns out to be a particular case of directed information helps to justify the former as a good measure of information flow, despite its symmetry: in channels without feedback it is a good measure *because it coincides with directed information* from input to output.

Lemma 9. *In absence of feedback,* $I(B^T \to A^T) = 0$

Proof. When feedback is not allowed, B^t and A_t are independent for $1 \leq t \leq T$. Then:

$$\begin{aligned}
& I(B^T \to A^T) \\
= & \sum_{t=1}^{T} I(A_t; B^{t-1}|A^{t-1}) && \text{(by Definition 2)} \\
= & \sum_{t=1}^{T} (H(A_t|A^{t-1}) - H(A_t|A^{t-1}, B^{t-1})) && \text{(by definition of mutual information)} \\
= & \sum_{t=1}^{T} (H(A_t|A^{t-1}) - H(A_t|A^{t-1})) && \text{(by the independence of } B^{t-1} \text{ and } A^t\text{)} \\
= & \ 0
\end{aligned}$$

□

Proposition 10. *In absence of feedback, leakage can be equivalently defined as directed information or as mutual information. Similarly, in absence of feedback, the maximum leakage can be equivalently defined as directed capacity or as capacity.*

Proof. It follows directly from Lemma 9 and (10). □

6 Modeling IIHSs as channels: An example

In this section we show the application of our approach to the *Cocaine Auction Protocol* [25]. The formalization of this protocol in terms of IIHSs using our framework makes it possible to prove the claim in [25] suggesting that if the seller knows the identity of the bidders then the (strong) anonymity guaranties are no longer assured.

Let us consider a scenario in which several mobsters 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 current bid value. This is done by adding an increment to the last bid. There

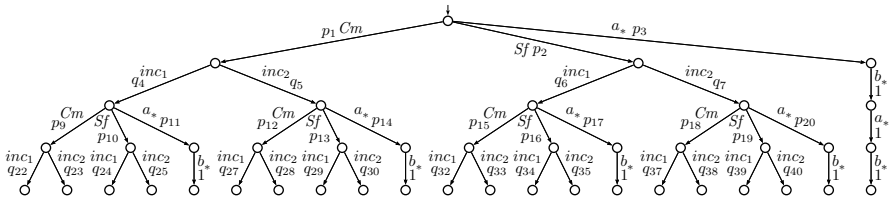


Figure 6: Cocaine Auction example

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 $\mathcal{I} = (M, \mathcal{A}, \mathcal{B})$, 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, and the probabilistic automaton M is represented in Figure 6. For clarity reasons, transitions with probability 0 are not represented in the automaton. Note that the special secret action a_* represents the situation where neither *Candlemaker* nor *Scarface* bid. The special observable action b_* represents the end of the auction and it can only occur if no one has bid in the round.

Table 5 shows all the stochastic kernels for this example.

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

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

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

$\alpha_1, \beta_1, \alpha_2 \to \beta_2$	inc_1	inc_2	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

Table 5: Stochastic kernels for the Cocaine Auction example.

The interested reader can find the construction of the reaction functions in Appendix C.

6.1 Calculating the information leakage

Let us now calculate the information leakage for this example using the concepts from Section 5. 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 mobsters 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 how the seller chooses the bid increment in 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 [25], where the seller does not have access to the identity of the bidder, for the sake of anonymity preservation. In general, we have $p(\beta_t|\alpha^t, \beta^{t-1}) = p(\beta_t|\beta^{t-1})$ for every $1 \leq t \leq T$. However, there is an exception: if there is no bidder, the case modeled by the secret being a_*, then the auction terminates, which is signaled by the observable b_*.

Example `b`: This is the most general case, without any restrictions. The presence of feedback allows the probability of the bidder to depend on the increment in the price. For instance, if *Candlemaker* is richer than *Scarface*, it is more likely that the former 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 as 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 as 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 up having *Scarface* as the final winner).

Example `c`: There is no feedback. In the cocaine auction, we can have the (perhaps 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 $p(\alpha_t|\alpha^{t-1}, \beta^{t-1}) = p(\alpha_t|\alpha^{t-1})$ for every $1 \leq t \leq T$. However, as in Example `b`, we do not impose any restriction on $p(\beta_t|\alpha^t, \beta^{t-1})$.

For each scenario we need to fill in the values of the probabilities in the protocol tree in Figure 6. The probabilities for each example are listed in Table 6. Table 7 shows a comparison between the values of the the entropy and of the directed information in the

Probability variable	Example `a` value	Example `b` value	Example `c` value
p_1	0.75	0.70	0.70
p_2	0.24	0.24	0.24
p_3	0.01	0.01	0.01
q_4	0.50	0.55	0.30
q_5	0.50	0.45	0.70
q_6	0.50	0.45	0.70
q_7	0.50	0.55	0.30
p_9	0.04	0.80	0.75
p_{10}	0.95	0.19	0.20
p_{11}	0.01	0.01	0.05
p_{12}	0.95	0.19	0.75
p_{13}	0.04	0.80	0.20
p_{14}	0.01	0.01	0.05
p_{15}	0.04	0.90	0.65
p_{16}	0.95	0.09	0.35
p_{17}	0.01	0.01	0.05
p_{18}	0.95	0.09	0.65
p_{19}	0.04	0.90	0.35
p_{20}	0.01	0.01	0.05
q_{22}	0.50	0.80	0.45
q_{23}	0.50	0.20	0.55
q_{24}	0.50	0.20	0.55
q_{25}	0.50	0.80	0.45
q_{27}	0.45	0.75	0.45
q_{28}	0.55	0.25	0.55
q_{29}	0.45	0.35	0.55
q_{30}	0.55	0.65	0.45
q_{32}	0.50	0.55	0.45
q_{33}	0.50	0.45	0.55
q_{34}	0.50	0.40	0.55
q_{35}	0.50	0.60	0.45
q_{37}	0.45	0.60	0.45
q_{38}	0.55	0.40	0.55
q_{39}	0.45	0.35	0.55
q_{40}	0.55	0.55	0.45

Table 6: Values of the probabilities in Figure 6 for Examples `a`, `b`, and `c`

three cases. 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., the leakage is low. The

Interpretation	Symbol	Example a	Example b	Example c
Input uncertainty	$H(A^T)$	1.9319	1.9054	1.9158
Reactor uncertainty	H_R	1.1911	1.5804	1.9158
A posteriori uncertainty	$H(A^T\|B^T)$	1.0303	1.2371	1.4183
Mutual information	$I(A^T;B^T) = H(A^T) - H(A^T\|B^T)$	0.9016	0.6684	0.4975
Leakage	$I(A^T \to B^T) = H_R - H(A^T\|B^T)$	0.1608	0.3433	0.4975
Feedback information	$I(B^T \to A^T)$	0.7408	0.3250	0.0000

Table 7: Values of the entropy and directed information for Examples a, b, and c

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 $I(A^T \to B^T)$.

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 represent the real leakage.

Finally, Figure 7 shows a comparison between the values of the entropy and of the directed information in the examples. The totality of the mutual information $I(A^T;B^T)$ is represented by the height of the correspondent bar, and we emphasize the contribution of the directed information in each direction by splitting the bar into two parts. This figure highlights the fact that mutual information can be misleading as a measure of leakage. The greatest mutual information is obtained in Example a, followed by Example b and then by Example c. However, the *real leakage*, given by $I(A^T \to B^T)$, respects exactly the inverse order, namely Example a presents the lowest value while Example c presents the highest one. Indeed, in Example a the value of $I(A^T \to B^T)$ represents only 18% of the mutual information, while in Example b it represents 51% and in Example c it amounts to 100%.

7 Topological properties of IIHSs and of their capacity

In this section we show how to extend to IIHSs the notion of pseudometric defined in [9] for Concurrent Labelled Markov Chains, and we prove that the capacity of the corresponding channels is a continuous function with respect to this pseudometric. The pseudometric construction is sound for general IIHSs, but the result on capacity is only valid for secret-nondeterministic IIHSs.

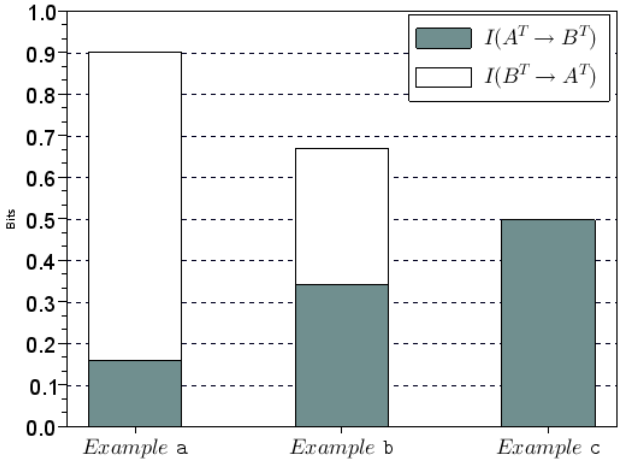


Figure 7: Comparison between the leakage in Examples a, b, and c

Given a set of states S, a pseudometric 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. We consider pseudometrics instead of metrics because our 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 pseudometrics 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 [31, 9, 8], we apply the construction based on the Kantorovich metric [14].

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 maximum is taken over 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. A pseudometric $d \in \mathcal{M}$ is a *bisimulation pseudometric*[2] 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 pseudometric is

$$d_{max} = \bigsqcup \{d \in \mathcal{M} \mid d \text{ is a bisimulation pseudometric}\} \tag{12}$$

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 pseudometric 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 [9, 8], which is characterized as follows (cf. Lemma 3.8 in the full version of [9], and Definition 2.7 in [8]):

$$F(d)(s, s') = \max\{\sup_{s \to \mu} \inf_{s' \to \mu'} d(\mu, \mu') \;,\; \sup_{s' \to \mu'} \inf_{s \to \mu} d(\mu, \mu')\}$$

Hence it can be proved, as an instance of the analogous result for F (cf. Lemma 2.8 in [8]), that $\Phi(d)$ is a pseudometric, and that the following property holds.

[2] In literature a pseudometric with this property is also called bisimulation *metric*, although it is a pseudometric, not a metric.

Lemma 11. *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 $d(\mu,\mu') \leq \epsilon$.*

From the above lemma and Definition 12 we derive (see also Lemma 2.9 in [8]):

Corollary 12. *A pseudometric d is a bisimulation pseudometric iff $d \preceq \Phi(d)$.*

By applying Corollary 12 to (12) we obtain

$$d_{max} = \bigsqcup \{d \in \mathcal{M} \mid d \preceq \Phi(d)\}$$

Furthermore, by adapting the proof of the monotonicity result for F (cf. Lemma 3.9 in the full version of [9]) we can prove the following:

Lemma 13. *Φ is monotonic on $(\mathcal{M} \preceq)$.*

Thanks to Lemma 13, and using Tarski's fixed point theorem as formulated in [29], we have that d_{max} is the greatest fixed point of Φ. Furthermore, by Corollary 12 we know that d_{max} is indeed a bisimulation pseudometric, and that it is the greatest bisimulation pseudometric.

In addition, the finite branching property of IIHSs ensures that the closure ordinal of Φ is ω (cf. Lemma 3.10 in the full version of [9]). Therefore we 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 14. *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 Φ, we have $d_{max} = \Phi(d_{max})$, and therefore

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

   which contradicts 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 $p(\beta_t \mid \alpha^t, \beta^{t-1}) \geq 0$. 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

$$d_{max}(\mathcal{I}(\alpha^{t-1}, \beta^{t-1}), \mathcal{I}'(\alpha^{t-1}, \beta^{t-1})) \geq d_{max}(\mathcal{I}(\alpha^t, \beta^{t-1}), \mathcal{I}'(\alpha^t, \beta^{t-1})). \tag{13}$$

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

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

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

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

which, together with (13), allows us to conclude. □

Lemma 15. *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 14(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

$$p(\beta_t \mid \alpha^t, \beta^{t-1}) - p'(\beta_t \mid \alpha^t, \beta^{t-1}) \leq d_{max}(\mathcal{I}(\alpha^{t-1}, \beta^{t-1}), \mathcal{I}'(\alpha^{t-1}, \beta^{t-1})). \tag{14}$$

By applying Lemma 14(2) $t - 1$ times, from (14) we get

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

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 pseudometric 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 pseudometric holds in general).

Theorem 16. *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 $T, \epsilon > 0$. Let $\mathcal{D}_T$ be the set of all input distributions in presence of feedback. Observe that

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

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 at 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 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 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 15, 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 [9] and we had to come up with a different reasoning.

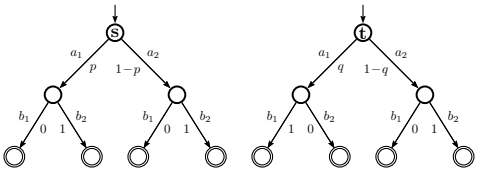
Example 6. Consider the two following programs, where a_1, a_2 are secrets, b_1, b_2 are observable, $\|$ is the parallel operator, and $+_p$ is a binary probabilistic choice that assigns probability p to the left branch, and probability $1-p$ to the right one.

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 8 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 classical 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.



(a)

s	b_1	b_2
a_1	0	1
a_2	0	1

(b)

t	b_1	b_2
a_1	1	0
a_2	0	1

Table 8: The IIHSs of Example 6 and corresponding channels, (a) for s and (b) for t.

8 Conclusion and discussion

In this paper we have investigated the problem of information leakage in interactive systems, and proved that these systems can be modeled as channels with memory and feedback. Furthermore, we have proved that the channel capacity is a continuous function of a pseudometric based on the Kantorovich metric.

We have considered various kinds of automata corresponding to different combinations of nondeterministic and probabilistic choice, as summarized in Table 9. Note that in this table the third row corresponds to the limit case in which the reactor is a Dirac measure, i.e. the probability is all concentrated on exactly one $\varphi^T \in \mathcal{F}$. It is easy to see that in this case $I(A^T \to B^T) = 0$ (all the entropies that constitute $I(A^T \to B^T)$ are 0), although $I(B^T \to A^T) \neq 0$. Therefore there is no leakage. In the classic case this corresponds to the situation in which the input distribution is a Dirac measure.

Table 10 summarizes the comparison between the channels with memory and feedback investigated in this paper, and the classic channels.

Throughout the paper we have assumed that the dependence of the secret choices on the observables is part of the external knowledge and, therefore, not considered

IIHSs **as automata**	IIHSs **as channels**	**Notion of leakage**
Normalized IIHSs with nondeterministic secrets and probabilistic observables	Sequence of stochastic kernels $\{p(\beta_t \vert \alpha^t, \beta^{t-1})\}_{t=1}^T$	Leakage as capacity
Fully probabilistic normalized IIHSs	Sequence of stochastic kernels $\{p(\beta_t \vert \alpha^t, \beta^{t-1})\}_{t=1}^T$ + reactor $\{p(\varphi_t \vert \varphi^{t-1})\}_{t=1}^T$	Leakage as directed information $I(A^T \to B^T)$
Normalized IIHSs with a deterministic scheduler solving the nondeterminism	Sequence of stochastic kernels $\{p(\beta_t \vert \alpha^t, \beta^{t-1})\}_{t=1}^T$ + sequence of reaction funct. φ^T	No leakage

Table 9: The various models considered in this paper

Classical channels	**Channels with memory and feedback**
The system is modeled in independent uses of the channel, often a unique use.	The system 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 $I(A^T \to B^T)$.

Table 10: Summary of the differences between classical and interactive systems

leakage. The reader may wonder what would happen if this assumption were dropped. We argue that in this case $I(B^T \to A^T)$ *could be considered as part of the leakage*. In the cases `a` and `b` of the cocaine auction example in Section 6, 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.

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 $I(B^T \to A^T)$ could represent a measure of the effectiveness of the adversary.

9 Future work

We would like to provide algorithms to compute the leakage and maximum leakage of interactive systems. These are rather challenging problems 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* [21] and *Stag and Hunt* [23] (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 and Frank Valencia for their useful remarks on early versions of this work.

References

[1] M. S. Alvim, M. E. Andrés, and C. Palamidessi. Information flow in interactive systems. In P. Gastin and F. Laroussinie, editors, *Proceedings of the 21st International Conference on Concurrency Theory (CONCUR 2010)*, volume 6269 of *Lecture Notes in Computer Science*, pages 102–116. Springer, 2010.

[2] 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.

[3] 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.

[4] K. Chatzikokolakis, C. Palamidessi, and P. Panangaden. Anonymity protocols as noisy channels. *Inf. and Comp.*, 206(2–4):378–401, 2008.

[5] D. Clark, S. Hunt, and P. Malacaria. Quantified interference for a while language. In *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.

[6] M. R. Clarkson, A. C. Myers, and F. B. Schneider. Belief in information flow. *Journal of Computer Security*, 17(5):655–701, 2009.

[7] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., second edition, 2006.

[8] 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.

[9] 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. Full version available at `http://www.cs.mcgill.ca/~prakash/Pubs/lics-full-sub.pdf`.

[10] Ebay website. `http://www.ebay.com/`.

[11] Ebid website. `http://www.ebid.net/`.

[12] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, New York, NY, 1968.

[13] J. W. Gray, III. Toward a mathematical foundation for information flow security. In *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy (SSP '91)*, pages 21–35, Washington - Brussels - Tokyo, May 1991. IEEE.

[14] 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.

[15] B. Köpf and D. A. Basin. An information-theoretic model for adaptive side-channel attacks. In P. Ning, S. D. C. di Vimercati, and P. F. Syverson, editors, *Proceedings of the 2007 ACM Conference on Computer and Communications Security, CCS 2007, Alexandria, Virginia, USA, October 28-31, 2007*, pages 286–296. ACM, 2007.

[16] 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.

[17] J. L. Massey. Causality, feedback and directed information. In *Proceedings of the 1990 International Symposium on Information Theory and its Applications*, November 1990.

[18] J. McLean. Security models and information flow. In *SSP'90*, pages 180–189. IEEE, 1990.

[19] Mercadolibre website. `http://www.mercadolibre.com/`.

[20] J. K. Millen. Hookup security for synchronous machines. In *Proceedings of the 3rd IEEE Computer Security Foundations Workshop (CSFW)*, pages 84–90, 1990.

[21] W. Poundstone. *Prisoners Dilemma*. Doubleday NY, 1992.

[22] R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, Massachusetts Institute of Technology, June 1995. Tech. Rep. MIT/LCS/TR-676.

[23] B. Skyrms. *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, 2003.

[24] 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.

[25] F. Stajano and R. J. Anderson. The cocaine auction protocol: On the power of anonymous broadcast. In *Information Hiding*, pages 434–447, 1999.

[26] W. Stallings. *Data and Computer Communications*. Prentice Hall, eigth edition, 2006.

[27] 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.

[28] A. Tanenbaum. *Computer Networks*. Prentice Hall, second edition, 1989.

[29] A. Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5(2):285—309, 1955.

[30] S. Tatikonda and S. K. Mitter. The capacity of channels with feedback. *IEEE Transactions on Information Theory*, 55(1):323–349, 2009.

[31] 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.

[32] W. Vickrey. Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance*, 16(1):8–37, 1961.

Appendix

A: An example illustrating the encoder/decoder design

In this section we consider again the erasure channel of Example 2 to show how the enriched model of channels with memory and feedback can be used to transmit the message, and in particular how the feedback can be used to design the encoder. We assume that the set $\mathcal{W}$ of possible messages consists of all finite sequences of bits. The role of the code functions is to encode the message W into a suitable representation for the stochastic kernels within the channel. The input alphabet for the stochastic kernels is $\mathcal{A} = \{0, 1\}$ and the output alphabet is $\mathcal{B} = \{0, 1, \mathtt{e}\}$, where the special output symbol $\mathtt{e}$ signals that a bit was erased. We assume that at most T uses of the channel are allowed and we use t, with $1 \leq t \leq T$, to represent the t^{th} time step.

We consider a sort of memory that depends only on the input history and we abstract from its specific form by defining a function $\mu : \wp_f(\mathcal{A}^t) \to [0, 1]$ that maps each possible input history to a correction factor to be added to (or subtracted from) a base probability value. We compute the contribution of μ to the base values using arithmetics modulo 2, in such a way that the resulting values are still a probability distribution. More precisely, the stochastic kernels are defined as follows.

$$\begin{array}{l} p(\beta_t = 0|\alpha^{t-1}0, \beta^{t-1}) = 0.8 - \mu(\alpha^{t-1}) \\ p(\beta_t = 1|\alpha^{t-1}0, \beta^{t-1}) = 0 \\ p(\beta_t = \mathtt{e}|\alpha^{t-1}0, \beta^{t-1}) = 0.2 + \mu(\alpha^{t-1}) \\ p(\beta_t = 0|\alpha^{t-1}1, \beta^{t-1}) = 0 \\ p(\beta_t = 1|\alpha^{t-1}1, \beta^{t-1}) = 0.8 - \mu(\alpha^{t-1}) \\ p(\beta_t = \mathtt{e}|\alpha^{t-1}1, \beta^{t-1}) = 0.2 + \mu(\alpha^{t-1}) \end{array} \tag{15}$$

Correspondingly, the general form of the channel matrix for each time $1 \leq t \leq T$ is shown in Table 11.

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

Table 11: General form of channel matrix for $1 \leq t \leq T$.

The code functions are chosen at time $t = 0$ based on the message to be transmitted. For illustration purposes, let us suppose that the message is the sequence of three bits $W = 011$. The other cases of W are analogous.

At time $t = 1$, the channel is used for its first time and the feedback history so far is empty $\beta^0 = \epsilon$. The encoder selects the input symbol $\alpha_0 = 0$, as in (16).

$$f_1[W = 011](\beta^0 = \epsilon) = 0 \tag{16}$$

At time $t = 2$, the feedback history consists of only one symbol, and in principle the possibilities are either $\beta^1 = 0$, $\beta^1 = 1$ or $\beta^1 = \mathtt{e}$. In the first case, the first bit was successfully transmitted and the encoder can go on to the second bit of the message. By the way the channel is defined, the second case is not really possible, so it is not important how the reaction function is defined for this case. We will denote this indifference by attributing to the function the symbol $\mathtt{x}$ instead of a 0 or a 1. In the last case, $\beta^1 = \mathtt{e}$, the first bit was erased and the encoder tries to retransmit the bit 0. We can write it formally as below.

$$\begin{array}{l} f_2[W = 011](\beta^1 = 0) = 1 \\ f_2[W = 011](\beta^1 = 1) = \mathtt{x} \\ f_2[W = 011](\beta^1 = \mathtt{e}) = 0 \end{array} \tag{17}$$

At time $t = 3$ the feedback histories allowed by the channel are $\beta^2 \in \{01, 0\mathtt{e}, \mathtt{e}0, \mathtt{ee}\}$ (the other ones have zero probability). In the first case, $\beta^2 = 01$ the two first bits of the message have been transmitted correctly and the encoder can send the third bit. If $\beta^2 = 0\mathtt{e}$, the transmission of the first bit was successful, but the second bit was erased and needs to be resent. In the case $\beta^2 = \mathtt{e}0$, the first bit was erased in the first try but was successfully transmitted in the second try, so now the encoder can move to the second bit of the message. In the last case, $\beta^2 = \mathtt{ee}$, the two tries were unsuccessful and the encoder still needs to transmit the first bit of the message. Formally:

$$\begin{array}{l} f_3[W = 011](\beta^2 = 00) = \mathtt{x} \\ f_3[W = 011](\beta^2 = 01) = 1 \\ f_3[W = 011](\beta^2 = 0\mathtt{e}) = 1 \\ f_3[W = 011](\beta^2 = 10) = \mathtt{x} \\ f_3[W = 011](\beta^2 = 11) = \mathtt{x} \\ f_3[W = 011](\beta^2 = 1\mathtt{e}) = \mathtt{x} \\ f_3[W = 011](\beta^2 = \mathtt{e}0) = 1 \\ f_3[W = 011](\beta^2 = \mathtt{e}1) = \mathtt{x} \\ f_3[W = 011](\beta^2 = \mathtt{ee}) = 0 \end{array} \tag{18}$$

We can easily extend the construction of code functions f_t for $3 \leq t \leq T$ using this encoding scheme.

The decoder is very simple: once all time steps $1, \ldots, T$ have taken place, it just takes the output whole trace β^T and removes the occurrences of the erased bit symbol $\mathtt{e}$ in order to recover the original message.

Table 12 shows a concrete example of a possible behavior of binary erasure channel with memory and feedback in a scenario where the message is $W = 011$ and the channel can be used $T = 3$ times. Note that in this particular example the maximum uses of the channel is achieved before the whole message is successfully sent: the decoder can recover only the two first bits of the original message.

B: Normalization of IIHS trees

In this section we address the problem of *normalizing* an IIHS, namely transforming it into a stratified automaton in which secret and observable actions alternate level by

Time t	Code functions $f_t(\beta^{t-1})$	Feedback history β^{t-1}	Encoder $\alpha_t = f_t[W](\beta^{t-1})$	Channel $p(\beta_t \mid \alpha^t, \beta^{t-1})$	Decoder $\hat{W} = \gamma(\beta^T)$
$t = 0$	Code functions for $W = 011$ are selected.	———	———	———	———
$t = 1$	As in (16)	ϵ	$\alpha_1 = f_1[W = 011](\epsilon) = 0$	According to $p(\beta_1 \mid 0, \epsilon)$ produces $\beta_1 = \mathtt{e}$	———
$t = 2$	As in (17)	$\mathtt{e}$	$\alpha_2 = f_2[W = 011](\mathtt{e}) = 0$	According to $p(\beta_2 \mid 00, \mathtt{e})$ produces $\beta_2 = 0$	———
$t = 3$	As in (18)	$\mathtt{e}0$	$\alpha_3 = f_3[W = 011](\mathtt{e}0) = 1$	According to $p(\beta_3 \mid 001, \mathtt{e}0)$ produces $\beta_3 = 1$	———
$t = 4$	———	———	———	———	Decoded message $\hat{W} = \gamma(\beta^3 = \mathtt{e}01) = 01$

Table 12: A possible evolution of the binary channel with time, for $W = 011$ and $T = 3$

level. The process of normalization described below is general enough to be applied to any IIHS without loss of generality or expressive power.

Let $\mathcal{A}$, $\mathcal{B}$ represent the secret ad observable actions, respectively. Consider a general IIHS $\mathcal{I} = (M, \mathcal{A}, \mathcal{B})$ with $M = (Q, \mathcal{L}, \hat{s}, \vartheta)$, where $\mathcal{L} = \mathcal{A} \cup \mathcal{B}$. Assume that we are interested only in executions that involve up to T interactions, i.e T uses of the system, with one secret taking place and one observable produced at each time.

In the normalization process, we unfold the automaton up to level $2T$, since there is one secret symbol and one observable symbol for each step. We also extend the secret alphabet $\mathcal{A}$ with a new symbol $a_* \notin \mathcal{A}$ and the observable alphabet $\mathcal{B}$ with a new symbol $b_* \notin \mathcal{B}$. These new symbols will be used as placeholders when we need to re-balance the tree. Let $\mathcal{A}' = \mathcal{A} \cup \{a_*\}$ and $\mathcal{B}' = \mathcal{B} \cup \{b_*\}$.

For a given level t let $\mathtt{Labels}(\mathcal{I}, t)$ be the set of all labels of transitions that can be performed with a non-zero probability from the states at the t^{th} level of the automaton. Formally:

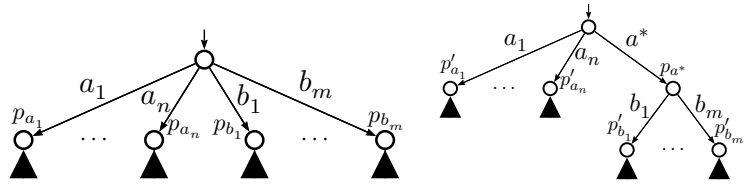
$$\mathtt{Labels}(\mathcal{I}, t) \equiv \{\ell \in \mathcal{L} \mid \exists \sigma, s \,.\, |\sigma| = t,\ \mathrm{last}(\sigma) \xrightarrow{\ell} s\}$$

The normalization of the IIHS $\mathcal{I}$ leads to an equivalent IIHS $\mathcal{I}' = (M', \mathcal{A}', \mathcal{B}')$, where $M' = (Q', \mathcal{L}', \hat{s}', \vartheta')$ and $\mathcal{L}' = \mathcal{A}' \cup \mathcal{B}'$; and such that, for every $1 \leq t \leq 2T$:

1. $\mathtt{Labels}(\mathcal{I}', t) \subseteq \mathcal{A}'$ or $\mathtt{Labels}(\mathcal{I}', t) \subseteq \mathcal{B}'$;
2. $\mathtt{Labels}(\mathcal{I}', t) \subseteq \mathcal{A}'$ iff $\mathtt{Labels}(\mathcal{I}', t+1) \subseteq \mathcal{B}'$, for $1 \leq t \leq T-1$;
3. $\mathtt{Labels}(\mathcal{I}', 1) \subseteq \mathcal{A}'$;

Condition 1 states that each level consists of either the secret actions only, or the observable actions only. Condition 2 states that secret and observable levels alternate. Condition 3 says that the automaton starts with a secret level.

First, the new symbols a_* and b_* are placeholders for the absence of a secret and observable symbol, respectively. If in a given level t we want to have only secret symbols, we can postpone the occurrences of observable symbols at this level as follows: add a_* to the secret level and "move" all the observable symbols to the subtree of a_*. Figure 8 exemplifies the local transformations we need to make on the tree.



(a) Local nodes of the tree before the transformation

(b) Local nodes of the tree after the transformation

Figure 8: Local transformation on an IIHS tree

Note that in Figure 8(b) the introduction of new nodes changed the probabilities of the transitions in the tree. In general, to normalize a secret level we need to introduce a_* in order to postpone the observable symbols, and the probabilities change as follows:

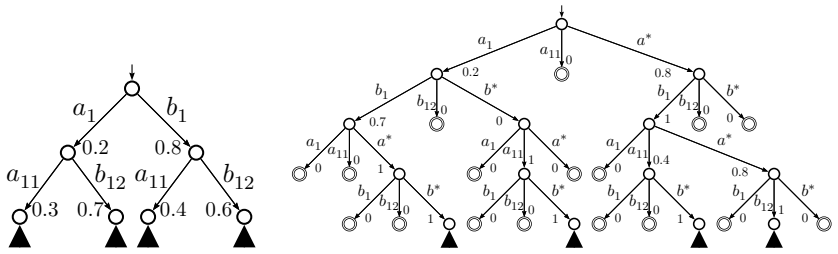
1. For every a_i, $1 \leq i \leq n$, the associated probability is maintained as $p'_{a_i} = p_{a_i}$;
2. The probability of the new symbol a_* is introduced as $p_{a_*} = \sum_{k=0}^{m} p_{b_k}$;
3. If $p_{a_*} \neq 0$, then for $1 \leq i \leq m$, the associated probability of b_j is updated to $p'_{b_j}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

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 subtree (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 normalizing an observable level, the same rules apply, guarding the proper symmetry between secrets and observables. We then proceed on the same way on the deeper levels of the tree. Figure 9 shows an example of a full transformation on a tree (for the sake of readability, we omit the levels where only $a_* = 1$ or $b_* = 1$).

C: Extended Cocaine Auction Protocol Example

We shall now extend the example of our approach applied to a real system. In Section 6 we introduced the Cocaine Auction Protocol and showed how to formalize one instance of it as an IIHS (Figure 6). We have also already defined the stochastic kernels for this example.



(a) Tree before transformation

(b) Tree after transformation

Figure 9: Transformation on an IIHS tree

The next step is to construct all the possible reaction functions $\{\varphi_t(\beta^{t-1})\}_{t=1}^{T}$. As seen in Section 4.2, 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. Table 13 contains the reaction functions for each time $t \leq 2$.

(a) All 3 reaction functions φ_1

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

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

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_*	-

Table 13: Reaction functions for the cocaine auction example.

Now we need to define the reactor, i.e., the probability distribution on reaction functions. Corollary 6 shows that we can do so by using the following equations:

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

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 $p(f_{2(6)}|f_{1(3)})$:

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

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

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