

Anonymity Protocols as Noisy Channels

Konstantinos Chatzikokolakis, Catuscia Palamidessi, Prakash Panangaden

▶ **To cite this version:**

Konstantinos Chatzikokolakis, Catuscia Palamidessi, Prakash Panangaden. Anonymity Protocols as Noisy Channels. Information and Computation, 2008, 206 (2-4), pp.378–401. 10.1016/j.ic.2007.07.003 . inria-00349225

HAL Id: inria-00349225
https://inria.hal.science/inria-00349225

Submitted on 25 Dec 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Anonymity Protocols as Noisy Channels ⋆

Konstantinos Chatzikokolakis [a],
Catuscia Palamidessi [a] Prakash Panangaden [b]

[a] *INRIA and LIX, École Polytechnique, Palaiseau, France*

[b] *School of Computer Science, McGill University, Montreal, Quebec, Canada*

Abstract

We consider a framework in which anonymity protocols are interpreted as noisy channels in the information-theoretic sense, and we explore the idea of using the notion of capacity as a measure of the loss of anonymity. Such idea was already suggested by Moskowitz, Newman and Syverson, in their analysis of the covert channel that can be created as a result of non-perfect anonymity. We consider the case in which some leak of information is intended by design, and we introduce the notion of conditional capacity to rule out this factor, thus retrieving a natural correspondence with the notion of anonymity. Furthermore, we show how to compute the capacity and the conditional capacity when the anonymity protocol satisfies certain symmetries. We also investigate how the adversary can test the system to try to infer the user's identity, and we study how his probability of success depends on the characteristics of the channel. We then illustrate how various notions of anonymity can be expressed in this framework, and show the relation with some definitions of probabilistic anonymity in literature. Finally, we show how to compute the matrix of the channel (and hence the capacity and conditional capacity) using model checking.

1 Introduction

In this paper we explore a general approach to measure the *degree of anonymity* provided by an anonymity protocol. Such protocols try to hide the link between

⋆ This work has been partially supported by the INRIA DREI Équipe Associée PRINTEMPS. The work of Konstantinos Chatzikokolakis and Catuscia Palamidessi has been also supported by the INRIA ARC project ProNoBiS.

Email addresses: kostas@lix.polytechnique.fr (Konstantinos Chatzikokolakis), catuscia@lix.polytechnique.fr (Catuscia Palamidessi), prakash@cs.mcgill.ca (Prakash Panangaden).

a set $\mathcal{A}$ of *anonymous events* and a set $\mathcal{O}$ of *observable events*. Events in $\mathcal{A}$ represent the information that we want to hide from the potential attacker. Ideally, we would like him to be totally unable to distinguish the events in $\mathcal{A}$, that is to deduce which of them really happened in a specific execution of the protocol. Events in $\mathcal{O}$ are the ones that the attacker actually observes. They should model all the possible outcomes of the protocol, from the point of view of the attacker. We assume that in each execution of the protocol one event $a \in \mathcal{A}$ and one event $o \in \mathcal{O}$ occur, and that o is disclosed to the attacker. An anonymity system should prevent the attacker from deducing a given the information about o and the knowledge about how the system works.

For example, a protocol could be designed to allow users to send messages to each other without revealing the identity of the sender. In this case, $\mathcal{A}$ would be the set of (the identities of) the possible users of the protocol, if only one user can send a message at a time, or the powerset of the users, otherwise. On the other hand, $\mathcal{O}$ could contain the sequences of all possible messages that the attacker can observe, depending on how the protocol works.

Probability plays an important role in anonymity protocols. First of all these protocols are very often probabilistic themselves. They use random primitives and the anonymity guarantees are based on the attacker's inability of determining the outcome of probabilistic choices. Clearly, the precise analysis of such protocols requires probabilistic means. Moreover, the analysis performed by the attacker can be also probabilistic, for example by gathering statistical information about the users. The attacker might not be able to find out exactly which anonymous event happened, but he could obtain a distribution over $\mathcal{A}$ and draw conclusions of the form "user i sent a message with probability 95%".

In this paper we consider a probabilistic setting, where probability distributions can be assigned to the elements of $\mathcal{A}, \mathcal{O}$. As a consequence we will model anonymous events by a random variable A on $\mathcal{A}$ and observable events by O on $\mathcal{O}$. From the point of view of the analysis, we are only interested in the distributions of A, O. In particular, the joint distribution $p(a, o)$ provides all the information about the conjoint behavior of the protocol and of the users that we need. From $p(a, o)$ we can derive, indeed, the marginal distributions $p(a)$ and $p(o)$, and the conditional distributions $p(o|a)$ and $p(a|o)$.

Most of the times, however, one is interested in abstracting from the specific set of users and its distribution, and proving properties about the protocol itself, aiming at *universal anonymity properties* that will hold no matter how the users behave (provided they follow the rules of the protocol). To this purpose, it is worth recalling that the joint distribution $p(a, o)$ can be decomposed as $p(a, o) = p(o|a)p(a)$. This decomposition singles out exactly the contributions of the protocol and of the users to the joint probability: $p(a)$, in fact, is the

probability associated to the users, while $p(o|a)$ represents the probability that the protocol produces o given that the users have produced a. The latter clearly depends only on the internal mechanisms of the protocol, not on the users.

This view of the protocol in isolation from the users brings us to consider the protocol as a device that, given $a \in \mathcal{A}$ as input, it produces an output in $\mathcal{O}$ according to a probability distribution $p(\cdot|a)$. This concept is well investigated in information theory, where such kind of device is called *channel*, and it is described by the matrix whose rows are the elements of $\mathcal{A}$, the columns the elements of $\mathcal{O}$, and the value in position (a, o) is the conditional probability $p(o|a)$. The rationale behind this view will be discussed in more details in Section 3.

1.1 Contribution

In this paper we investigate the idea of measuring the degree of anonymity of a protocol in terms of the information-theoretic notion of *capacity* of the protocol, seen as channel. Our original contribution consist of the following:

- We define a more general notion of capacity, that we call *conditional capacity*, which models the case in which some loss of anonymity is allowed by design.
- We discuss how to compute the capacity and the conditional capacity when the anonymity protocol satisfies certain symmetries.
- We investigate the relation between the channel's matrix and the knowledge that an attacker can gain on the anonymous actions (the channel's inputs) from the observables (the channel's outputs). In particular, we consider attackers following the Bayesian approach to *hypothesis testing*, and we show bounds on the probability of error (also known as Bayesian risk) regarding the probabilistic information that the attacker can acquire.
- We compare the definition of with various probabilistic notions of anonymity given in literature, in particular perfect anonymity, relative anonymity, and probable innocence. Finally, we show that the condition of probable innocence corresponds to a certain information-theoretic bound.
- We show how to compute the matrix of a protocol using model checking tools. We demonstrate our ideas in the dining cryptographers and Crowds protocols, where we show how the parameters of each protocol affect its anonymity.

Several various formal definitions and frameworks for reasoning about anonymity have been developed in literature. These include approaches based on process-calculi [1,2], epistemic logic [3,4], and "function views" [5].

Several methods and protocols to guarantee anonymity have also been proposed. They are based on very diverse techniques, depending on the application domain. For instance, in a recent work that has attracted much attention [6], a notion called k-anonymity is satisfied if the information relative to each person in a data release is indistinguishable from the one of at least other $k-1$ individuals in the same release. Such situation can be achieved by generalizing some fields in the records and by suppressing others [7].

In this paper, we focus on protocols that use randomized mechanisms to achieve anonymity. In such context it is natural to explore probabilistic and information-theoretic approaches.

Probabilistic definitions of anonymity have been investigated in [8,4,9–11]. We discuss the relation with these works in detail in Section 5.

A recent line of work has been dedicated to exploring the notion of anonymity from an information-theoretic point of view [12,13]. The main difference with our approach is that in those works the anonymity degree is expressed in terms of entropy, rather than mutual information. More precisely, the emphasis is on the lack of information that an attacker has about the distribution of the users, rather than on the capability of the protocol to conceal this information despite of the observables that are made available to the attacker. Moreover, a uniform user distribution is assumed, while in this paper we try to abstract from the user distribution and make no assumptions about it.

Channel capacity has been already used in an anonymity context in [14,15], where the ability to have covert communication as a result of non-perfect anonymity is examined. The difference with our approach is that in those works the channels are constructed by the users of the protocol using the protocol mechanisms, to the purpose of transferring information, and capacity is used to measure the amount of information that can be transferred through these channels. In our paper, we consider the channel to be an abstraction of the protocol itself, and we use the capacity to measure the anonymity degree of the protocol. However in [15] the authors also suggest that the channel's capacity can be used as an asymptotic measure of the worst-case loss of anonymity, which is the idea that we explore in this paper. Note that in [15] the authors warn that in certain cases the notion of capacity might be too strong a measure to compare systems with, because the holes in the anonymity of a system might not behave like text book discrete memoryless channels.

Zhu and Bettati propose in [16] a definition of anonymity based on mutual information. The notion we consider is based on capacity, which is an abstraction of mutual information obtained by maximizing over the possible input distributions. As a consequence, we get a measure that depends only on the protocol (i.e. the channel) and not on the users (i.e. the input distribution), which is an advantage because in general we don't know the input distribution, and it also depend on the users, and even with the same users, it may change over time. Of course, in case we know a priori the input distribution, then the definition of Zhu and Bettati is more precise because it gives the exact loss of anonymity for the specific situation.

A different information-theoretic approach is taken in [17]. In this paper, the authors define as information leakage the difference between the a priori accuracy of the guess of the attacker, and the a posteriori one, after the attacker has made his observation. The accuracy of the guess is defined as the Kullback-Leibler distance between the *belief* (which is a weight attributed by the attacker to each input hypothesis) and the true distribution on the hypotheses.

Another approach close in spirit to ours is the one of [18]. In this work, the authors use the Kullback-Leibler distance to perform a metric analysis of anonymity. In our work, we use the notion of mutual information, which is a special case of relative entropy. However, the specific application of relative entropy in [18] is radically different from ours. We use it to compare the entropy of the input of an anonymity protocol before and after the observation. They use it to establish a sort of distance between the traces of an anonymity system.

In the field of information flow and non-interference there is a line of research which is closely related to ours. There have been various works [19–23] in which the *high information* and the *low information* are seen as the input and output respectively of a channel. From an abstract point of view, the setting is very similar; technically it does not matter what kind of information we are trying to conceal, what is relevant for the analysis is only the probabilistic relation between the input and the output information. The conceptual and technical novelties of this paper w.r.t. the above works are explained in Section 1.1. We believe that part of our framework and of our results are applicable more or less directly also to the field of non-interference. Some of the results however, for instance those based on the hypotheses of symmetry or weak symmetry of the protocol's matrix, seem to be specific of the anonymity setting, in the sense that the assumptions would be too restrictive for the non-interference case.

The relation between the adversary's goal of inferring a secret from the observables, and the field of "hypothesis testing", has been explored in other papers in literature, see in particular [24–26]. To our knowledge, however, this is the

first time that it is investigated in connection with the matrix of conditional probabilities determined by the protocol.

1.3 Plan of the paper

Next section recalls some basic notions about information theory. In Section 3 we justify our view of protocols as channels and (loss of) anonymity as capacity and conditional capacity, and we give a method to compute these quantities in special symmetry cases. In Section 4 we consider the tests that an attacker can make on the protocol in order to gain knowledge about the anonymous actions, and we discuss the probability of error that limits the inferences based on such tests. In Section 5, we relate our framework to other probabilistic approaches to anonymity. Finally, in Section 6, we illustrate on two specific examples (the dining cryptographers and Crowds) how to compute the channel matrix and the degree of anonymity for a given protocol, possibly using automated tools.

2 Preliminaries on Information Theory

Being in a purely probabilistic setting gives us the ability to use tools from information theory to reason about the uncertainty of a random variable and the information that it can reveal about another random variable. In particular the notions we will be interested in are *entropy*, *mutual information* and *channel capacity*. In this section we briefly revise these notions. We refer to [27] for more details.

In general, we will use capital letters X, Y to denote random variables and the corresponding calligraphic letters $\mathcal{X}, \mathcal{Y}$ for their set of values. We will also use small letters x, y to represent values of these variables, $p(x), p(y)$ to denote the probability of x and y respectively and $p(x, y)$ to denote the joint probability of x and y.

Let X be a random variable. The *entropy* $H(X)$ of X is defined as $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$. The entropy measures the uncertainty of a random variable. It takes its maximum value $\log |\mathcal{X}|$ when X's distribution is uniform and its minimum value 0 when X is constant. We usually take the logarithm with a base 2 and measure entropy in *bits*. Roughly speaking, m bits of entropy means that we have 2^m values to choose from, assuming a uniform distribution.

The *relative entropy* or *Kullback–Leibler distance* between two probability distributions p, q on the same set $\mathcal{X}$ is defined as $D(p \parallel q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$. It is possible to prove that $D(p \parallel q)$ is always non-negative, and it is 0 if and

only if $p = q$.

Now let X, Y be random variables. The *conditional entropy* $H(X|Y)$ is $H(X|Y) = -\sum_{y\in\mathcal{Y}} p(y) \sum_{x\in\mathcal{X}} p(x|y) \log p(x|y)$. Conditional entropy measures the amount of uncertainty of X when Y is known. It can be shown that $0 \leq H(X|Y) \leq H(X)$. It takes its maximum value $H(X)$ when Y reveals no information about X, and its minimum value 0 when Y completely determines the value of X.

Comparing $H(X)$ and $H(X|Y)$ gives us the concept of *mutual information* $I(X;Y)$, which is defined as $I(X;Y) = H(X) - H(X|Y)$. Mutual information measures the amount of information that one random variable contains about another random variable. In other words, it measures the amount of uncertainty about X that we lose when observing Y. It can be shown that it is symmetric ($I(X;Y) = I(Y;X)$) and that $0 \leq I(X;Y) \leq H(X)$.

A *communication channel* is a tuple $\langle \mathcal{X}, \mathcal{Y}, p(\cdot|\cdot)\rangle$ where $\mathcal{X}, \mathcal{Y}$ are the sets of input and output symbols respectively and $p(y|x)$ is the probability of observing output $y \in \mathcal{Y}$ when $x \in \mathcal{X}$ is the input. Given an input distribution $p(x)$ over $\mathcal{X}$ we can define the random variables X, Y for input and output respectively. The maximum mutual information between X and Y over all possible distributions $p(x)$ is known as the channel's *capacity*: $C = \max_{p(x)} I(X;Y)$. The capacity of a channel gives the maximum rate at which information can be transmitted using this channel.

3 Loss of Anonymity as Channel Capacity

The notions discussed in previous section can be used to reason about the information that the adversary obtains from the protocol. The entropy $H(A)$ of A gives the amount of uncertainty about the anonymous events, before executing the protocol. The higher the entropy is the less certain we are about the outcome of A. After the execution, however, we also know the actual value of O. Thus, the conditional entropy $H(A|O)$ gives the uncertainty of the attacker about the anonymous events after performing the observation. To compare these two entropies, we consider the mutual information $I(A;O)$ which measures the information about A that is contained in O. This measure is exactly what we want to minimize. In the best case it is 0, meaning that we can learn nothing about A by observing O (in other words $H(A|O)$ is equal to $H(A)$). In the worst case it is equal to $H(A)$ meaning that all the uncertainty about A is lost after the observation, thus we can completely deduce the value of A ($H(A|O)$ is 0).

As explained in the introduction, each execution of an anonymity protocol is associated to the joint probability $p(a, o)$ of the particular values taken by A, O

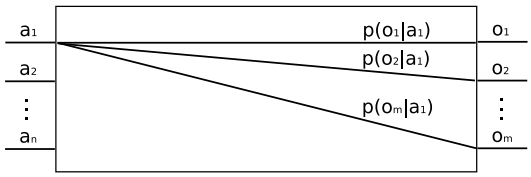


Fig. 1. An anonymity channel

in that execution. This probability can be written as $p(a, o) = p(a)p(o|a)$. In our view, among these two values, $p(o|a)$ can be considered as a characteristic of the protocol, while $p(a)$ depends only on the users. For instance, in a protocol for sender anonymity, A takes values on the set $\mathcal{A}$ of users, and $p(a)$ is the probability of user a being the sender. In some cases all users might have the same probability of being the sender, in other cases a particular user might send messages more often than the others. Since the design of the protocol should be independent from the particular users who will use it, the analysis of the protocol should make no assumptions about the distribution on $\mathcal{A}$. On the other hand $p(o|a)$ gives the probability of o when a is the sender, so it depends only on the internal mechanisms of the protocol, not on how often a sends messages.

To abstract from the probabilities of the anonymous events, we view an anonymity protocol as a channel $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$ where the sets of anonymous events $\mathcal{A}$ and observable events $\mathcal{O}$ are the input and output alphabets respectively, and the matrix $p(o|a)$ gives the probability of observing o when a is the input. An anonymity channel is shown in Figure 1. Different distributions of the input will give different values of $I(A; O)$. We are interested in the worst possible case, so we adopt the definition of the *loss of anonymity* as the maximum value of $I(A; O)$ over all possible input distributions, that is the capacity of the corresponding channel. We recall that this idea was already suggested in [15].

Definition 1 *Let $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$ be an anonymity protocol. The* loss of anonymity *C of the protocol is defined as*

$$C = \max_{p(a)} I(A; O)$$

where the maximum is taken over all possible input distributions.

The loss of anonymity measures the amount of information about A that can be learned by observing O in the worst possible distribution of anonymous events. If it is 0 then, no matter what is the distribution of A, the attacker can learn nothing more by observing the protocol. In fact, as we will see in section 5.1, this corresponds exactly to notions of perfect anonymity in literature [8,4,9]. However, as we discuss in section 5.3, our framework also captures weaker notions of anonymity.



Fig. 2. A simple elections protocol

As with entropy, channel capacity is measured in bits. Roughly speaking, 1 bit of capacity means that after the observation A will have one bit less of entropy, in another words the attacker will have reduced the set of possible users by a factor 2, assuming a uniform distribution.

3.1 Relative Anonymity

So far, we have assumed that ideally no information about the anonymous events should be leaked. However, there are cases where *some* information about the anonymous events is allowed to be revealed by design, without this leak be considered as a flaw of the protocol. Consider, for example, the case of a simple elections protocol, displayed in figure 2. For simplicity we assume that there are only two candidates c and d, and that each user always votes for one of them, so an anonymous event can be represented by the subset of users who voted for candidate c. In other words, $\mathcal{A} = 2^V$ where V is the set of voters. The output of the protocol is the list of votes of all users, however, in order to achieve anonymity, the list is randomly reordered, using for example some MIX technique[1] . As a consequence, the attacker can see the number of votes for each candidate, although he should not be able to find out who voted for whom. Indeed, determining the number of votes of candidate c (the cardinality of a), while concealing the vote expressed by each individual (the elements that constitute a), is the purpose of the protocol.

So it is clear that after the observation only a fraction of the anonymous events remains possible. Every event $a \in \mathcal{A}$ with $|a| \neq n$ where n is the number of votes for candidate c can be ruled out. As a consequence $H(A|O)$ will be smaller than $H(A)$ and the capacity of the corresponding channel will be non-zero, meaning that some anonymity is lost. In addition, there might be a loss of anonymity due to other factors, for instance, if the reordering technique is not uniform. However, it is undesirable to confuse these two kinds of anonymity losses, since the first is by design and thus acceptable. We would like a notion of anonymity that factors out the *intended* loss and measures only the loss that we want to minimize.

[1] In MIX protocols an agent waits until it has received requests from multiple users and then forwards the requests in random order to hide the link between the sender and the receiver of each request.

In order to cope with the intended anonymity loss, we introduce a random variable R whose outcome is the revealed information. In the example of the elections protocol, the value of R is the cardinality of a. Since we allow to reveal R by design, we can consider that R is known even before executing the protocol. So, $H(A|R)$ gives the uncertainty about A given that we know R and $H(A|R,O)$ gives the uncertainty after the execution of the protocol, when we know both R and O. By comparing the two we retrieve the notion of *conditional mutual information* $I(A;O|R)$ defined as

$$I(A;O|R) = H(A|R) - H(A|R,O)$$

So, $I(A;O|R)$ is the amount of uncertainty on A that we lose by observing O, given that R is known. Now we can define the notion of *conditional capacity* $C|R$ which will give us the *relative loss of anonymity* of a protocol.

Definition 2 *Let $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$ be an anonymity protocol and R a random variable defined by its set of values $\mathcal{R}$ and a probability matrix $p(r|a,o)$. The* relative loss of anonymity *of the protocol with respect to R is defined as*

$$C|R = \max_{p(a)} I(A;O|R)$$

where the maximum is taken over all possible input distributions.

3.1.1 Partitions: a special case of relative anonymity

An interesting special case of relative anonymity is when the knowledge of either an anonymous event or an observable event totally determines the value of R. In other words, both $\mathcal{A}$ and $\mathcal{O}$ are partitioned in subsets, one for each possible value of R. The elections protocol of the previous section is an example of this case. In this protocol, the value r of R is the number of votes for candidate A. This is totally determined by both anonymous events a (r is the cardinality of a) and observable events o (r is the number of c's in o). So we can partition $\mathcal{A}$ in subsets $\mathcal{A}_0, \ldots, \mathcal{A}_n$ such that $|a| = n$ for each $a \in \mathcal{A}_n$, and similarly for $\mathcal{O}$. Notice that an anonymous event $a \in \mathcal{A}_i$ produces only observables in $\mathcal{O}_i$, and vice versa.

In this section we show that such systems can be viewed as the composition of smaller, independent sub-systems, one for each value of R.

We say that R is a deterministic function of X if $p(r|x)$ is 0 or 1 for all $r \in \mathcal{R}$ and $x \in \mathcal{X}$. In this case we can partition $\mathcal{X}$ as follows

$$\mathcal{X}_r = \{x \in \mathcal{X} \mid p(r|x) = 1\}$$

Clearly the above sets are disjoint and their union is $\mathcal{X}$.

Theorem 3 *Let $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$ be an anonymity protocol and R a random variable defined by its set of values $\mathcal{R} = \{r_1, \ldots, r_l\}$ and a probability matrix $p(r|a,o)$. If R is a deterministic function of both A and O, under some non-zero input distribution $p(\cdot)$ [2], then the transition matrix of the protocol is of the form*

	$\mathcal{O}_{r_1}$	$\mathcal{O}_{r_2}$	$\cdots$	$\mathcal{O}_{r_l}$
$\mathcal{A}_{r_1}$	M_{r_1}	0	$\ldots$	0
$\mathcal{A}_{r_2}$	0	M_{r_2}	$\ldots$	0
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\mathcal{A}_{r_l}$	0	0	$\ldots$	M_{r_l}

and

$$C|R \leq d \quad \Leftrightarrow \quad C_i \leq d, \forall i \in 1..l$$

where C_i is the capacity of the channel with matrix M_{r_i}.

PROOF. First we show that the protocol matrix has the above form, that is $p(o|a) = 0$ if $a \in \mathcal{A}_r, o \in \mathcal{O}_{r'}$ with $r \neq r'$. If $p(o) = 0$ then (since $p(\cdot)$ is non-zero) then whole column of o is zero and we are finished. Otherwise, since R is a deterministic function of A, O we have $p(r|a) = 1$ and $p(r|o) = 0$. Then

$$p(r, a|o) = 0 \Rightarrow p(r, o|a)\frac{p(a)}{p(o)} = 0 \Rightarrow p(r, o|a) = 0$$

Finally

$$p(r \vee o|a) = p(r|a) + p(o|a) - p(r, o|a) = 1 + p(o|a)$$

so $p(o|a) = 0$ otherwise $p(r \vee o|a)$ would be greater than 1.

Now we show that $C|R \leq d$ iff $C_i \leq d, \forall i \in 1..l$ where C_i is the capacity of the channel with matrix M_{r_i}, constructed by taking only the rows in $\mathcal{A}_{r_i}$ and the columns in $\mathcal{O}_{r_i}$.

($\Rightarrow$) Assume that $C|R \leq d$ but $\exists i : C_i > d$. Then there exists a distribution p_i over $\mathcal{A}_{r_i}$ such that $I(A_{r_i}; O_{r_i}) > d$ where A_{r_i}, O_{r_i} are the input and output random variables of channel M_{r_i}. We construct a distribution over $\mathcal{A}$ as follows

$$p(a) = \begin{cases} p_i(a) \text{ if } a \in \mathcal{A}_{r_i} \\ 0 \qquad \text{otherwise} \end{cases}$$

[2] We require $p(\cdot)$ to assign non-zero probability to all users so that $p(r|o)$ can be defined unless the whole column is zero. Note that if R is a deterministic function of O under some non-zero distribution, it is also under all distributions.

It is easy to see that under that distribution, $I(A;O|R) = I(A_{r_i}|O_{r_i})$ which is a contradiction since $I(A;O|R) \leq C|R \leq d < I(A_{r_i}|O_{r_i})$.

($\Leftarrow$) The idea is that for each input distribution $p(a)$ we can construct an input distribution $p_r(a)$ for each sub-channel M_r and express $I(A;O|R)$ in terms of the mutual information of all sub-channels. We write $I(A;O|R)$ as:

$$\begin{aligned}
&I(A;O|R)\\
&= H(A|R) - H(A|R,O)\\
&= -\sum_{r\in\mathcal{R}} p(r) \sum_{a\in\mathcal{A}} p(a|r)\log p(a|r) + \sum_{\substack{r\in\mathcal{R}\\ o\in\mathcal{O}}} p(r,o) \sum_{a\in\mathcal{A}} p(a|r,o)\log p(a|r,o)\\
&= -\sum_{r\in\mathcal{R}} p(r)\Big[\sum_{a\in\mathcal{A}} p(a|r)\log p(a|r) - \sum_{o\in\mathcal{O}} p(o|r) \sum_{a\in\mathcal{A}} p(a|r,o)\log p(a|r,o)\Big]
\end{aligned}$$

Moreover, we have

$$p(a|r) = \begin{cases} \dfrac{p(a)}{p(r)} & \text{if } a \in \mathcal{A}_r \\ 0 & \text{otherwise} \end{cases} \qquad p(o|r) = \begin{cases} \dfrac{p(o)}{p(r)} & \text{if } o \in \mathcal{O}_r \\ 0 & \text{otherwise} \end{cases}$$

Also $p(a|r,o) = p(a|o)$ if $o \in \mathcal{O}_r$ and $p(a|r,o) = 0$ if $a \notin \mathcal{A}_r$. Thus in the above sums the values that do not correspond to each r can be eliminated and the rest can be simplified as follows:

$$I(A;O|R) = -\sum_{r\in\mathcal{R}} p(r)\Big[\sum_{a\in\mathcal{A}_r} \frac{p(a)}{p(r)}\log\frac{p(a)}{p(r)} - \sum_{o\in\mathcal{O}_r} \frac{p(o)}{p(r)} \sum_{a\in\mathcal{A}_r} p(a|o)\log p(a|o)\Big] \tag{1}$$

Now for each $r \in \mathcal{R}$ we define a distribution p_r over $\mathcal{A}_r$ as follows:

$$p_r(a) = \frac{p(a)}{p(r)}$$

It is easy to verify that this is indeed a probability distribution. We use p_r as the input distribution in channel M_r and since, by construction of M_r, $p_r(o|a) = p(o|a)$ we have

$$p_r(o) = \sum_{a\in\mathcal{A}_r} p_r(a)p_r(a|o) = \sum_{a\in\mathcal{A}_r} \frac{p(a)}{p(r)} p(a|o) = \frac{p(o)}{p(r)}$$

Now equation (1) can be written:

$$\begin{aligned}
&I(A;O|R)\\
&\quad= \sum_{r\in\mathcal{R}} p(r)\Big[-\sum_{a\in\mathcal{A}_r} p_r(a)\log p_r(a) + \sum_{o\in\mathcal{O}_r} p_r(o) \sum_{a\in\mathcal{A}_r} p_r(a|o)\log p_r(a|o)\Big]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{r \in \mathcal{R}} p(r) \Big[H(A_r) - H(A_r|O_r) \Big] \\
&= \sum_{r \in \mathcal{R}} p(r) I(A_r; O_r) \\
&\leq \sum_{r \in \mathcal{R}} p(r) d \\
&= d
\end{aligned}
$$

Where A_r, O_r are the input and output random variables of channel M_r. Finally, since $I(A; O|R) \leq d$ for all input distributions we have $C|R \leq d$. □

3.2 Computing the channel's capacity

For arbitrary channels, there is no analytic formula to compute their capacity. In the general case we can only use numerical algorithms that converge to the capacity, as we discuss in the end of this section. In practice, however, channels have symmetry properties that can be exploited to compute the capacity in an easy way. In this section we define classes of symmetry and discuss how to compute the capacity for each class. Two classic cases are the *symmetric* and *weakly symmetric* channels.

Definition 4 *A matrix is* symmetric *if all rows are permutations of each other and all columns are also permutations of each other. A matrix is* weakly *symmetric if all rows are permutations of each other and the column sums are equal.*

The following result is from literature:

Theorem 5 ([27], page 189) *Let $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot) \rangle$ be a channel. If $p(\cdot|\cdot)$ is weakly symmetric then the channel's capacity is given by a uniform input distribution and is equal to*

$$C = \log |\mathcal{O}| - H(\mathbf{r})$$

where $\mathbf{r}$ is a row of the matrix and $H(\mathbf{r})$ is the entropy of $\mathbf{r}$.

Note that symmetric channels are also weakly symmetric so Theorem 5 holds for both classes.

In anonymity protocols, users usually execute exactly the same protocol, with the only difference being the names of the agents to whom they communicate. So if a user a_1 produces an observable o_1 with probability p, it is reasonable to assume that a_2 will produce some observable o_2 with the same probability. In other words we expect all rows of the protocol's matrix to be permutations of each other. On the other hand, the columns are not necessarily permutations of each other, as we will see in the example of Section 6. The problem is that o_1

and o_2 above need not be necessarily different. We can have observables that are produced with equal probability by all users. Clearly, these "constant" columns cannot be the permutation of non-constant ones so the resulting channel matrix will not be symmetric (and not even weakly symmetric).

To cope with this kind of channels we define a more relaxed kind of symmetry called *partial symmetry*. In this class we allow some columns to be constant and we require the sub-matrix, composed only by the non-constant columns, to be symmetric. A weak version of this symmetry can also be defined.

Definition 6 *A matrix is* partially symmetric *(resp. weakly partially symmetric) if some columns are constant (possibly with different values in each column) and the rest of the matrix is symmetric (resp. weakly symmetric).*

Now we can extend Theorem 5 to the case of partial symmetry.

Theorem 7 *Let $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$ be a channel. If $p(\cdot|\cdot)$ is weakly partially symmetric then the channel's capacity is given by*

$$C = p_s \log \frac{|\mathcal{O}_s|}{p_s} - H(\mathbf{r}_s)$$

where $\mathcal{O}_s$ is the set of symmetric output values, $\mathbf{r}_s$ is the symmetric part of a row of the matrix and p_s is the sum of $\mathbf{r}_s$.

PROOF. Let $\mathcal{O}_s$ by the set of symmetric output values (the ones that correspond to the symmetric columns) and $\mathcal{O}_n$ the set of the non-symmetric ones. Also let $\mathbf{r}$ be a row of the matrix and $\mathbf{r}_s$ the symmetric part of $\mathbf{r}$. Since the matrix is partially symmetric all rows are permutations of each other. As a consequence:

$$H(O|A) = -\sum_o p(o) \sum_a p(o|a) \log p(o|a) = H(\mathbf{r})$$

Moreover the columns in $\mathcal{O}_n$ are constant so for all $o \in \mathcal{O}_n$, $p(o)$ is independent of the input distribution: $p(o) = \sum_a p(a)p(o|a) = p(o|a')$ for some fixed a'. We have

$$\begin{aligned} I(A;O) &= H(O) - H(O|A) \\ &= -\sum_{o\in\mathcal{O}} p(o)\log p(o) - H(\mathbf{r}) \\ &= -\sum_{o\in\mathcal{O}_s} p(o)\log p(o) - \sum_{o\in\mathcal{O}_n} p(o|a')\log p(o|a') - H(\mathbf{r}) \\ &= -\sum_{o\in\mathcal{O}_s} p(o)\log p(o) - H(\mathbf{r}_s) \end{aligned}$$

$$\leq -\sum_{o\in\mathcal{O}_s} \frac{p_s}{|\mathcal{O}_s|}\log\frac{p_s}{|\mathcal{O}_s|} - H(\mathbf{r}_s) \tag{2}$$

$$= p_s \log\frac{|\mathcal{O}_s|}{p_s} - H(\mathbf{r}_s) \tag{3}$$

We constructed inequality (2) by taking a uniform distribution $p(o) = \frac{p_s}{|\mathcal{O}_s|}$ of symmetric outputs (the non-symmetric outputs have constant probabilities). p_s is the total probability of having an output among those in $\mathcal{O}_s$. Now if we take a uniform input distribution $p(a) = \frac{1}{|\mathcal{A}|}$ then for all $o \in \mathcal{O}_s : p(o) = \sum_a p(a)p(o|a) = \frac{c}{|\mathcal{A}|}$ where c is the sum of the corresponding column which is the same for all symmetric output values. So a uniform input distribution produces a uniform distribution of the symmetric output values, thus the bound (3) is achieved and it is the actual capacity of the channel. □

Note that Theorem 7 is a generalization of Theorem 5. A (weakly) symmetric channel can be considered as (weakly) partially symmetric with no constant columns. In this case $O_s = O, \mathbf{r}_s = \mathbf{r}, p_s = 1$ and we retrieve Theorem 5 from Theorem 7.

In all cases of symmetry discussed above, computing the capacity is a simple operation involving only one row of the matrix and can be performed in $O(|\mathcal{O}|)$ time.

In the general case of no symmetry we must use a numerical algorithm, like the Arimoto-Blahut algorithm (see for instance [27]) which can compute the capacity to any desired accuracy. However the convergence rate is slow (linear) and the coefficient of the convergence speed gets smaller when the number of input values increases.

4 Testing anonymous events

In this section we illustrate the relation between the channel's matrix and the possibility for the attacker of guessing the anonymous event from the consequent observable event. This problem is known in statistics literature as *hypothesis testing*. The idea is that we have a set of data or outcomes of an experiment, and a set of possible alternative explanations (*hypotheses*). We have to infer which hypothesis holds from the data, possibly by repeating the experiment, and try to minimize the probability of guessing the wrong hypothesis (*probability of error*).

We assume that the same hypothesis holds through the repetition of the experiment, which means that the same user is re-executing the protocol multiple

times, either forced by the attacker himself or by some external factor. For instance, in Crowds ([10]) users send messages along randomly selected routes. For various reasons this path might become unavailable, so the user will need to create an new one, thus re-executing the protocol. If the attacker is part of the path, he could also cause it to fail by stop forwarding messages, thus obliging the sender to recreate it (unless measures are taken to prevent this, as it is done in Crowds).

We also assume that the random variables corresponding to the outcomes of the experiments are independent. This corresponds to assuming that the protocol is memoryless, i.e. each time it is reactivated, it works according to the same probability distribution, independently from what happened in previous sessions.

In statistics there are several frameworks and methods for hypothesis testing. We consider here the Bayesian approach, which requires the knowledge of the matrix of the protocol and of the *a priori* distribution of the hypotheses, and tries to infer the *a posteriori* probability of the actual hypothesis w.r.t. a given observation or sequence of observations. The first assumption (knowledge of the matrix of the protocol) is usually granted in an anonymity setting, since the way the protocol works is public. The second assumption may look too strong, since the attacker does not usually know the distribution of the anonymous actions. We show, however, that under certain conditions the a priori distribution becomes less and less relevant with the repetition of the experiment, and, at the limit, it does not matter at all.

Let us introduce some notation. Given an anonymous event a, consider the situation in which the user re-executes the protocol n times with the same a as input event, and the attacker tries to infer a from the n observable outputs of the protocol executions. Let $O_1, O_2, \ldots, O_n$ represent the random variables corresponding to the observations made by the attacker, and let $\vec{o}$ denote a sequence of observed outputs $o_1, o_2, \ldots o_n$. As stated above, we assume that O_1, O_2, $\ldots$, O_n are independent, hence the distribution of each of them is given by $p(\cdot|a)$, and their conjoint distribution $p : \mathcal{O}^n \to [0,1]$ is given by

$$p(\vec{o}|a) = \prod_{i=1}^{n} p(o_i|a) \tag{4}$$

Let $f_n : \mathcal{O}^n \to \mathcal{A}$ be the *decision function* adopted by the adversary to infer the anonymous action from the sequence of observables. Let $E_{f_n} : \mathcal{A} \to \mathcal{O}^n$ be the function that gives the *error region* of f_n when $a \in \mathcal{A}$ has occurred, namely:

$$E_{f_n}(a) = \{\vec{o} \in \mathcal{O}^n \mid f_n(\vec{o}) \neq a\}$$

Finally, let $\eta_n : \mathcal{A} \to [0,1]$ be the function that associates to each $a \in \mathcal{A}$ the probability of inferring the wrong input event on the basis of f_n when $a \in \mathcal{A}$

has occurred, namely:

$$\eta_n(a) = \sum_{\vec{o} \in E_{f_n}(a)} p(\vec{o}|a)$$

We are now ready to introduce the *probability of error* associated to anonymous action testing on a given anonymity protocol, following the lines of the Bayesian approach (see for instance [27], Section 12.8).

Definition 8 *Given an anonymity protocol $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot) \rangle$, a sequence of n experiments, and a decision function f_n, the* probability of error P_{f_n} *is defined as the probability weighted sum over $\mathcal{A}$ of the individual probabilities of error. Namely:*

$$P_{f_n} = \sum_{a \in A} p(a)\eta_n(a)$$

In the Bayesian framework, the best possible decision function is given by the so-called *maximum a posteriori rule*, which, given the sequence of observables $\vec{o} \in \mathcal{O}^n$, tries to maximize the a posteriori probability of the hypothesis a w.r.t. $\vec{o}$. The a posteriori probability of a w.r.t. $\vec{o}$ is given by Bayes theorem (aka Bayes Inversion Rule):

$$p(a|\vec{o}) = \frac{p(\vec{o}|a)p(a)}{p(\vec{o})}$$

We now define a class of decision functions based on the above approach.

Definition 9 *Given an anonymity protocol $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot) \rangle$, and a sequence of n experiments, a decision function f_n is a* Bayesian decision function *if for each $\vec{o} \in \mathcal{O}^n$, $f_n(\vec{o}) = a$ implies $p(\vec{o}|a)p(a) \geq p(\vec{o}|a')p(a')$ for every $a' \in \mathcal{A}$.*

The above definition is justified by the following result which is a straightforward consequence of known results in literature.

Proposition 10 *Given an anonymity protocol $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot) \rangle$, a sequence of n experiments, and a Bayesian decision function f_n, for any other decision function h_n we have that $P_{f_n} \leq P_{h_n}$.*

PROOF. Immediate from the fact that the maximum a posteriori rule minimizes the probability of error. See, for instance, [27], Section 12. □

4.1 Independence from the input distribution

The definition of the Bayesian decision functions depends on the a priori probability distribution on $\mathcal{A}$. This might look artificial, since in general such distribution is unknown. We will show, however, that under a certain condition

on the matrix of the protocol, for n large enough, the Bayesian decision functions and the associated probability of error do not depend on the distribution on $\mathcal{A}$.

The following definition establishes the condition on the matrix.

Definition 11 *Given an anonymity protocol $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$, we say that such protocol is determinate iff all rows are pairwise different, i.e. the probability distributions $p(\cdot|a)$, $p(\cdot|a')$ are different for each pair a, a' with $a \neq a'$.*

Next proposition shows that if a protocol is determinate, then it can be approximated by a decision function which compares only the elements along the column corresponding to the observed event, without considering the input probabilities. By "approximated" we mean that as n increases, the probability of the subset of $\mathcal{O}^n$ in which the two functions give the same result converges to 1.

This property is based on a remark in [27], page 316, stating that, for n large enough, in the fraction $p(\vec{o}|a)p(a)/p(\vec{o}|a')p(a')$ the factor $p(a)/p(a')$ is dominated by the factor $p(\vec{o}|a)/p(\vec{o}|a')$ (provided, one needs to add, that the latter is different from 1). In [27] they give also a sketch of the proof of this remark; the proof of our proposition is is a development of that sketch.

Proposition 12 *Given a determinate anonymity protocol $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$, for any distribution $p(\cdot)$ on $\mathcal{A}$, any Bayesian decision functions f_n, and any decision function $g_n : \mathcal{O}^n \to \mathcal{A}$ such that $g_n(\vec{o}) = a$ implies $p(\vec{o}|a) \geq p(\vec{o}|a')$ for all $a' \in \mathcal{A}$, we have that g_n approximates f_n. Namely, for any $\epsilon > 0$, there exists n such that the probability of the set $\{\vec{o} \in \mathcal{O}^n \mid f_n(\vec{o}) \neq g_n(\vec{o})\}$ is smaller than ϵ.*

PROOF. For any value $o \in \mathcal{O}$, and for any sequence of observable outcomes $\vec{o} \in \mathcal{O}^n$, let $n(o, \vec{o})$ denote the number of o's that occur in $\vec{o}$. Let a be the actual input. Observe that, by the strong law of large numbers ([27]), for any $\delta > 0$ the probability of the set $\{\vec{o} \in \mathcal{O}^n \mid \forall o \in \mathcal{O}\ |n(o, \vec{o})/n - p(o|a)| < \delta\}$ goes to 1 as n goes to ∞. We show that, as a consequence of the above observation, the probability of the set $S = \{\vec{o} \in \mathcal{O}^n \mid \forall a' \neq a\ p(\vec{o}|a)p(a) > p(\vec{o}|a')p(a')\}$ goes to 1 as n goes to ∞. In fact, $p(\vec{o}|a)p(a) > p(\vec{o}|a')p(a')$ iff

$$\frac{1}{n}\log\frac{p(\vec{o}|a)p(a)}{p(\vec{o}|a')p(a')} > 0$$

and

$$\frac{1}{n}\log\frac{p(\vec{o}|a)p(a)}{p(\vec{o}|a')p(a')} = \frac{1}{n}\log\frac{p(\vec{o}|a)}{p(\vec{o}|a')} + \frac{1}{n}\log\frac{p(a)}{p(a')}$$

$$
\begin{aligned}
&\xrightarrow[n\to\infty]{} \frac{1}{n}\log\frac{p(\vec{o}|a)}{p(\vec{o}|a')} && \text{(since } \frac{1}{n}\log\frac{p(a)}{p(a')} \xrightarrow[n\to\infty]{} 0\text{)}\\
&= \frac{1}{n}\log\prod_{i=1}^{n}\frac{p(o_i|a)}{p(o_i|a')} && \text{(by (4))}\\
&= \frac{1}{n}\sum_{i=1}^{n}\log\frac{p(o_i|a)}{p(o_i|a')}\\
&= \frac{1}{n}\sum_{o\in\mathcal{O}} n(o,\vec{o})\log\frac{p(o|a)}{p(o|a')} && \text{(by definition of } n(o,\vec{o})\text{)}\\
&\xrightarrow[n\to\infty]{} \sum_{o\in\mathcal{O}} p(o|a)\log\frac{p(o|a)}{p(o|a')} && \text{(strong law of large numb.)}\\
&= D(p(\cdot|a) \parallel p(\cdot|a')) && \text{(Kullback–Leibler distance)}\\
&> 0 && \text{(by determinacy)}
\end{aligned}
$$

Given a Bayesian decision function f_n, consider now the set $S' = \{\vec{o} \in \mathcal{O}^n \mid f_n(\vec{o}) = a\}$. Because of the definition of f_n, we have that $S \subseteq S'$. Hence also the probability of the set S' goes to 1 as n goes to ∞. Following a similar reasoning, we can prove that for any g_n satisfying the premises of proposition, the probability of the set $\{\vec{o} \in \mathcal{O}^n \mid g_n(\vec{o}) = a\}$ goes to 1 as n goes to ∞. We can therefore conclude that the same holds for the probability of the set $\{\vec{o} \in \mathcal{O}^n \mid g_n(\vec{o}) = f_n(\vec{o})\}$. □

Proposition 12 allows us to define a decision function, for n sufficiently large, by comparing only the probabilities $p(\vec{o}|a)$ for different a's. These probabilities are determined uniquely by the matrix and therefore no knowledge of the a priori probability on A is required.

The conditional probability $p(o|a)$ (resp. $p(\vec{o}|a)$) is called *likelihood* of a given o (resp. $\vec{o}$). The criterion for the definition of g_n used in Proposition 12 is to choose the a which maximizes the likelihood of o, and it is known in literature as the *maximum likelihood rule*.

4.2 Bounds on the probability of error

In this section we discuss some particular cases of matrices and the corresponding bounds on the error that can be introduced by the Bayesian decision functions. Some more cases will be considered in the next section.

We start with the bad case (from the anonymity point of view), which is when the matrix is determinate:

Proposition 13 *Given a determinate anonymity protocol $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$, for any distribution $p(\cdot)$ on $\mathcal{A}$, and for any ϵ, there exists n such that the property*

$$g_n(\vec{o}) = a \text{ implies } p(\vec{o}|a) \geq p(\vec{o}|a') \text{ for all } a' \in \mathcal{A}$$

determines a unique decision function g_n on a set of probability greater than $1-\epsilon$, and the probability of error P_{g_n} is smaller than ϵ.

PROOF. Given $\vec{o} \in \mathcal{O}^n$, define $g_n(\vec{o}) = a$ iff a is the value of $\mathcal{A}$ for which $p(\vec{o}|a)$ is greatest. By following the same lines as in the proof of Proposition 12, we have that the set $\{\vec{o} \in \mathcal{O}^n \mid \forall a' \in \mathcal{A}\ p(\vec{o}|a) > p(\vec{o}|a')\}$ has probability greater than $1-\epsilon$ for n sufficiently large. Consequently, the choice of a is unique.

As for P_{g_n}, we observe that for n sufficiently large the set $E_{g_n} = \{\vec{o} \in \mathcal{O}^n \mid \exists a' \in \mathcal{A}\ p(\vec{o}|a) \leq p(\vec{o}|a')\}$ has probability smaller than ϵ. Hence $\eta_n(a) = \sum_{\vec{o} \in E_{g_n}(a)} p(\vec{o}|a) < \epsilon$ and $P_{g_n} = \sum_{a \in \mathcal{A}} p(a)\eta_n(a) < \epsilon$. □

Proposition 13 and its proof tell us that, in case of determinate matrices, there is essentially only one decision function, and its value is determined, for n sufficiently large, by the a for which $p(\vec{o}|a)$ is greatest.

Consider now the opposite case, i.e. when there are at least two identical rows in the matrix, in correspondence of a_1 and a_2. In such case, for the sequences $\vec{o} \in \mathcal{O}^n$ such that $p(\vec{o}|a_1)(= p(\vec{o}|a_2))$ is maximal, the value of g_n is not uniquely determined, because we could choose either a_1 or a_2. Assuming that we choose arbitrarily between them, and that the probability of choosing the wrong one is uniformly distributed, we have that the probability of error is bound from below as follows [3]: $P_{g_n} = \sum_{a \in \mathcal{A}} p(a)\eta_n(a) \geq p(a_1)1/2 + p(a_2)1/2$.

More in general, if there are k identical rows $a_1, a_2, \ldots, a_k$, the lower bound to the probability of error is $P_{g_n} = \sum_{a \in \mathcal{A}} p(a)\eta_n(a) \geq p(a_1)(k-1)/k + p(a_2)(k-1)/k + \ldots + p(a_k)(k-1)/k$.

The situation is slightly different if we know the a priori distribution and we define the function f_n. In this case, the criterion of maximizing $p(a)p(\vec{o}|a)$ reduces to maximizing $p(a)$. Hence, observing the outcome of the protocol does not add any information to what we already know. However, the a priori knowledge can help to make a sensible guess about the most likely a. This is

[3] Note that this bound is strict. In fact, using the strong law of large numbers it is possible to prove that, when either a_1 or a_2 is the actual input, the probability of the set of the sequences $\vec{o} \in \mathcal{O}^n$ for which $p(\vec{o}|a_1)$ (and $p(\vec{o}|a_2)$) is maximal goes to 1 as n goes to ∞.

not the case, of course, if in addition to rows a_1 and a_2 being identical we also have $p(a_1) = p(a_2)$.

5 Relation with existing anonymity notions

In this section we consider some particular channels, and we illustrate the relation with probabilistic (non information-theoretic) notions of anonymity existing in literature.

5.1 Capacity 0*: strong anonymity*

The case in which the capacity of the anonymity protocol is 0 is by definition obtained when $I(A;O) = 0$ for all possible input distributions of $\mathcal{A}$. From information theory we know that this is the case iff A and O are independent (cfr. [27], page 27). Hence we have the following characterization:

Proposition 14 *Given an anonymity system $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$, the capacity of the corresponding channel is* 0 *iff all the rows of the channel matrix are the same, i.e. $p(o|a) = p(o|a')$ for all o, a, a'.*

The condition $p(o|a) = p(o|a')$ *for all* o, a, a' has been called *strong probabilistic anonymity* in [9] and it is equivalent to the condition $p(a|o) = p(a)$ *for all* o, a. The latter was considered as a definition of anonymity in [8] and it is called *conditional anonymity* in [4].

Capacity 0 is the optimal case, of course, also w.r.t. the capability of the adversary of testing the anonymous events (cfr. Section 4): All the rows are the same, hence $p(\vec{o}|a_1) = p(\vec{o}|a_2)$ for all $a_1, a_2 \in \mathcal{A}$, and $\vec{o} \in \mathcal{O}^n$. Consequently the observations are of no use for the attacker to infer the anonymous event, i.e. to define the "right" $g_n(\vec{o})$, since all $p(\vec{o}|a)$ are maximal. Assuming a uniform distribution in assigning a value to $g_n(\vec{o})$, the probability of error is bound from below by $(|\mathcal{A}| - 1)/|\mathcal{A}|$ (cfr. Section 4.2).

An example of protocol with capacity 0 is the *dining cryptographers* in a connected graph [8], under the assumption that it is always one of the cryptographers who pays, and that the coins are fair.

5.2 Conditional capacity 0: strong anonymity "within a group"

In some anonymity protocols, the users are divided in groups and the protocol allows the adversary to figure out to which group the culprit belongs, although it tries to conceal which user in the group is the culprit. This is the case, for example, of the dining cryptographers in a generic graph [8], where the groups correspond to the connected components of the graph.

Such situation corresponds to having a partition on $\mathcal{A}$ and $\mathcal{O}$, see Section 3.1. The case of conditional capacity 0 is obtained when each M_{r_i} has capacity 0, namely when in each group r_i the rows are identical.

From the point of view of testing the anonymous events we note the following: given a $\vec{o} \in \mathcal{O}^n$, there exists exactly one group r_i of a's such that $p(\vec{o}|a) > 0$, and $p(\vec{o}|a_1) = p(\vec{o}|a_2)$ for all a_1, a_2 in r_i. Hence the attacker knows that the "right" value of $g_n(\vec{o})$ is an a in r_i, but he does not know exactly which one. In other words, on the basis of the observations the attacker can get complete knowledge about the group, but remains completely uncertain about the exact event a in the group, as expected. The lower bound on the probability of error is $(|\mathcal{A}_r| - 1)/|\mathcal{A}_r|$ where $r \in \mathcal{R}$ determines the set of maximal cardinality in $\mathcal{A}$.

Under the assumption that the coins are fair it can be shown that the dining cryptographers in a generic graph has conditional capacity 0 ([8]).

One of the authors of [28], David Sands, has suggested to us that the notion of strong anonymity "within a group" seems related to the notion of equivalence classes in his work. Exploring this connection is left for future work.

5.3 Probable innocence: weaker bounds on capacity

Probable innocence is a weak notion of anonymity introduced by Reiter and Rubin for the Crowds protocol [10]. In this section we focus on the definition of probable innocence, a description and analysis of Crowds can be found in Section 6.2.

Probable innocence was verbally defined as "from the attacker's point of view, the sender appears no more likely to be the originator of the message than to not be the originator". In literature there are three different definitions [10,4,11] that try to formally express this notion, see [11] for details. In this section we discuss the relation between these definitions and the channel capacity.

5.3.1 Definition of Reiter and Rubin

In [10] Reiter and Rubin gave a verbal definition of probable innocence and then formalized it and proved it for the Crowds protocol. Their formalization considers the probability that the originator forwards a message directly to a corrupted member (the attacker) and requires this probability to be at most one half. The event of forwarding a message to the attacker is an observable event: the attacker can detect it during the execution of the protocol. Thus, Reiter and Rubin's formalization of probable innocence considers the probability of observable events produced by user a_i when he executes the protocol [4]. As explained in [11], this definition could be expressed in the framework of this paper as follows: a protocol satisfies RR-probable innocence iff [5]

$$p(o|a) \leq \frac{1}{2} \; \forall o \in \mathcal{O}, \forall a \in \mathcal{A} \tag{5}$$

In [11] it is argued that this definition makes sense for Crowds due to certain properties that Crowds satisfies, however it is not suitable for arbitrary protocols.

We now show that RR-probable innocence imposes no bound on the capacity of the corresponding channel. Consider, for example, the protocol shown in figure 3. The protocols satisfies RR-probable innocence since all values of the matrix are less than or equal to one half. However the channel capacity is (the matrix is symmetric) $C = \log|\mathcal{O}| - H(\mathbf{r}) = \log(2n) - \log 2 = \log n$ which is the maximum possible capacity, equal to the entropy of A. Indeed, users can be perfectly identified by the output since each observable is produced by exactly one user.

Note, however, that in Crowds a bound on the capacity can be obtained due to the special symmetries that it satisfies which make RR-probable equivalent to CP-probable innocence.

5.3.2 Definition of Halpern and O'Neill

In [4] Halpern and O'Neill give a definition of probable innocence that focuses on the attacker's confidence that a particular anonymous event happened, after performing an observation. It requires that the probability of an anonymous

[4] Note that this probability has little to do with the probability of a_i to actually execute the protocol. The latter can be arbitrarily small or big.

[5] Note that in [11] this definition is given as $p(o|a) \leq \frac{1}{2}p(h)$ where $p(h)$ is the probability that the message passes at least once from the attacker. To simplify the analysis we consider only the cases that this happens, in other words $p(h) = 1$. This consideration is orthogonal to our discussion, the same result can be obtained without it.

	o_1	o_2	o_3	o_4	$\cdots$	o_{2n-1}	o_{2n}
a_1	1/2	1/2	0	0	...	0	0
a_1	0	0	1/2	1/2	...	0	0
$\vdots$	$\vdots$				$\ddots$		$\vdots$
a_n	0	0	0	0	...	1/2	1/2

Fig. 3. A maximum-capacity channel which satisfies RR-probable innocence

event should be at most one half, under any observation. A protocol satisfies HO-probable innocence iff

$$p(a|o) \leq \frac{1}{2} \; \forall o \in \mathcal{O}, \forall a \in \mathcal{A} \tag{6}$$

This definition looks like the one of Reiter and Rubin but its meaning is very different. It does not limit the probability of observing o. Instead, it limits the probability of an anonymous event a given the observation of o.

As discussed in [11], the problem with this definition is that it depends on the probabilities of the anonymous events which are not part of the protocol. As a consequence, HO-probable innocence cannot hold for all input distributions. If we consider a distribution where $p(a)$ is very close to 1, then $p(a|o)$ cannot possibly be less than 1/2. So we cannot speak about the bound that HO-probable innocence imposes to the capacity, since to compute the capacity we quantify over all possible input distributions and HO-probable innocence cannot hold for all of them. However, if we limit ourselves to the input distributions where HO-probable innocence actually holds, then we can prove the following proposition.

Proposition 15 *Let $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot) \rangle$ be a channel and $p(a)$ a fixed distribution over $\mathcal{A}$. If the channel is symmetric and satisfies HO-probable innocence for this input distribution then $I(A;O) \leq H(A) - 1$.*

PROOF. If X is a random variable and f a function on $\mathcal{X}$, we will denote by $Ef(X)$ the expected value of $f(X)$. Note that $H(X) = -E \log p(X)$ and $H(X|Y) = -E \log p(X|Y)$.

We have

$$I(A;O) = H(A) - H(A|O) = H(A) + E \log p(A|O)$$

And since $p(A|O) \leq 1/2$ and both log and E are monotonic

$$I(A;O) \leq H(A) + E \log \frac{1}{2} = H(A) - 1$$

□

Note that we consider the mutual information for a specific input distribution, not the capacity, for the reasons explained above.

5.3.3 Definition of Chatzikokolakis and Palamidessi

The definition of [11] tries to combine the other two by considering both the probability of producing some observable and the attacker's confidence after the observation. This definition considers the probability of two anonymous events a, a' producing the same observable o and does not allow $p(o|a)$ to be too high or too low compared to $p(o|a')$. A protocol satisfies CP-probable innocence iff

$$(n-1)p(o|a') \geq p(o|a) \quad \forall o \in \mathcal{O}, \forall a, a' \in \mathcal{A} \tag{7}$$

where $n = |\mathcal{A}|$. In [11] it is shown that this definition overcomes some drawbacks of the other two definitions of probable innocence and it is argued that it is more suitable for general protocols. In this section we show that CP-probable innocence imposes a bound on the capacity of the corresponding channel, which strengthens our belief that it is a good definition of anonymity.

Since the purpose of this definition is to limit the fraction $\frac{p(o|a)}{p(o|a')}$ we could generalize it by requiring this fraction to be less than or equal to a constant γ.

Definition 16 *An anonymity protocol* $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot) \rangle$ *satisfies* partial anonymity *if there is a constant* γ *such that*

$$\gamma\, p(o|a') \geq p(o|a) \quad \forall o \in \mathcal{O}, \forall a, a' \in \mathcal{A}$$

A similar notion is called *weak probabilistic anonymity* in [29].

Note that partial anonymity generalizes both CP-probable innocence ($\gamma = n-1$) and strong probabilistic anonymity ($\gamma = 1$). The following theorem shows that partial anonymity imposes a bound to the channel capacity:

Theorem 17 *Let* $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot) \rangle$ *be an anonymity protocol. If the protocol is symmetric and satisfies partial anonymity with* $\gamma > 1$ *then*

$$C \leq \frac{\log \gamma}{\gamma - 1} - \log \frac{\log \gamma}{\gamma - 1} - \log \ln 2 - \frac{1}{\ln 2}$$

PROOF. Since the channel is symmetric, by Theorem 5 its capacity is given by $\log|\mathcal{O}| - H(\mathbf{r})$ where $\mathbf{r}$ is a row of the matrix. We consider the first row which contains values of the form $p(o|a_1), o \in \mathcal{O}$. Since the columns are permutations of each other, we have $\forall o \exists a : p(o|a_1) = p(o_1|a)$. And since the protocol satisfies partial anonymity we have $\forall a, a' \in \mathcal{A} : \gamma\, p(o_1|a') \geq p(o_1|a)$, thus

$$\gamma\, p(o'|a_1) \geq p(o|a_1) \qquad \forall o, o' \in \mathcal{O} \tag{8}$$

First we show that when we decrease the distance between the probabilities in a distribution then the entropy increases (this is a standard result from information theory). Let $\vec{x} = (x_1, x_2, \ldots, x_n)$ such that $x_1 < x_2$ and let $\vec{x_o} = (x_1 + d, x_2 - d, \ldots, x_n)$ with $d \leq x_2 - x_1$. We can write $\vec{x_o}$ as a convex combination $t\vec{x} + (1-t)\vec{x_p}$ where $t = 1 - \frac{d}{x_2 - x_1}$ and $\vec{x_p} = (x_2, x_1, \ldots, x_n)$. Since $H(\vec{x}) = H(\vec{x_p})$ and $H(\vec{x})$ is a concave function of $\vec{x}$ ([27]) we have

$$H(\vec{x_o}) = H(t\vec{x} + (1-t)\vec{x_p}) \geq tH(\vec{x}) + (1-t)H(\vec{x_p}) = H(\vec{x})$$

Let p be the minimum value of the row $\mathbf{r}$. By (8) the maximum value of $\mathbf{r}$ will be at most γp. To maximize the capacity we want to minimize $H(\mathbf{r})$ so we will construct the row which gives the minimum possible entropy without violating (8). If there are any values of the row between p and γp we could subtract some probability from one and add it to another value. Since this operation increases the distance between the values, it decreases the entropy of the row as we showed before (in the inverse direction). So for a fixed p the lowest entropy is given by the row whose values are either p or γp. After that we can no longer separate the values without violating (8). However, this is a local optimum. If we take a new p' and construct a new row with values p' and $\gamma p'$ then we might find an even lower entropy.

Let x be the number of elements with value γp. Also let $m = |\mathcal{O}|$. We have

$$(m - x)p + x\gamma p = 1 \Rightarrow p = \frac{1}{A} \quad \text{with} \quad A = x(\gamma - 1) + m$$

And the entropy of $\mathbf{r}$ will be

$$\begin{aligned} H(\mathbf{r}) = h(x) &= -(m-x)\frac{1}{A}\log\frac{1}{A} - x\frac{\gamma}{A}\log\frac{\gamma}{A} \\ &= (-x(\gamma-1) - m)\frac{1}{A}\log\frac{1}{A} - x\frac{\gamma}{A}\log\gamma \\ &= \log A - x\frac{\gamma}{A}\log\gamma \end{aligned}$$

So $H(\mathbf{r})$ is a function $h(x)$ of only one variable x. We want to find the value x_0 which minimizes $h(x)$. Note that x_0 could be fractional, meaning that we cannot split exactly the row into p_0 and γp_0 elements. In this case $h(x_0)$ will

not correspond to an achievable probability of a row, but it will still be a lower bound. First we derive $h(x)$

$$h'(x) = \frac{1}{\ln 2}\frac{\gamma - 1}{A} - \gamma \log \gamma \frac{m}{A^2}$$

And x_0 will be the value for which

$$\begin{aligned} h(x_0) = 0 \Rightarrow \\ \frac{1}{\ln 2}\frac{\gamma - 1}{x_0(\gamma - 1) + m} = \frac{m\gamma \log \gamma}{(x_0(\gamma - 1) + m)^2} \Rightarrow \\ x_0 = \frac{A_0 - m}{\gamma - 1} \quad \text{with} \\ A_0 = \frac{m\gamma \log \gamma \ln 2}{\gamma - 1} \end{aligned}$$

Finally the minimum entropy of $\mathbf{r}$ will be equal to

$$\begin{aligned} h(x_0) &= \log \frac{m\gamma \log \gamma \ln 2}{\gamma - 1} - \frac{\gamma \log \gamma}{\gamma - 1} + \frac{1}{\ln 2} \\ &= \log m - \frac{\log \gamma}{\gamma - 1} + \log \log \gamma - \log(\gamma - 1) + \log \ln 2 + \frac{1}{\ln 2} \end{aligned}$$

And the maximum capacity will be

$$\begin{aligned} C_{\max} &= \log m - h(x_0) \\ &= \frac{\log \gamma}{\gamma - 1} - \log \frac{\log \gamma}{\gamma - 1} - \log \ln 2 - \frac{1}{\ln 2} \end{aligned}$$

□

This bound has two interesting properties. First, it depends only on γ and not on the number of input or output values or on other properties of the channel matrix. Second, the bound converges to 0 as $\gamma \to 1$. As a consequence, due to the continuity of the capacity as a function of the channel matrix, we can retrieve Proposition 14 about strong probabilistic anonymity ($\gamma = 1$) from Theorem 17. A bound for probable innocence can be obtained by taking $\gamma = n - 1$, so Theorem 17 treats strong anonymity and probable innocence in a uniform way. Note that this bound is proved for the special case of symmetric channels, we plan to examine the general case in the future.

Concerning the testing of the anonymous events, it is interesting to note that, if the attacker has the possibility of repeating the test with the same input an arbitrary number of times, then probable innocence does not give any guarantee. In fact, condition 7 does not prevent the function $p(\vec{o}|\cdot)$ from having a maximum with probability close to 1, for a sufficiently long sequence of observables $\vec{o}$. So we can define $g_n(\vec{o})$ to be such maximum, and we have that

the probability of error corresponding to g_n goes to 0. The only exception is when two (or more) rows a_1, a_2 are equal and correspond to maximals. Imposing this condition for all anonymous actions is equivalent to requiring strong anonymity. In conclusion, probable innocence maintains an upper bound on anonymity through protocol repetition only if the system is strongly anonymous. This result is in accordance to Proposition 17 in [11].

6 Computing the degree of anonymity of a protocol

In this section we discuss how to compute the channel matrix and the degree of anonymity for a given protocol, possibly using automated tools. We illustrate our ideas on two protocols from literature: the dining cryptographers ([8]), and Crowds ([10]), while, at the same time, we try to convey some general heuristic principles.

6.1 Dining cryptographers

6.1.1 Description of the protocol

This protocol, proposed by Chaum in [8], is arguably the most known anonymity protocol in the literature. The protocol is usually demonstrated in a situation where three cryptographers are dining together with their master. At the end of the dinner, each of them is secretly informed by the master whether he should pay the bill or not. So, either the master will pay, or he will ask one of the cryptographers to pay. The cryptographers, or some external observer, would like to find out whether the payer is one of them or the master. However, if the payer is one of them, they also wish to maintain anonymity over the identity of the payer. Of course, we assume that the master himself will not reveal this information, and also we want the solution to be distributed, i.e. communication can be achieved only via message passing, and there is no central memory or central *coordinator* which can be used to find out this information.

The Dining Cryptographers protocol offers a solution to this problem. Each cryptographer tosses a coin which is visible to himself and to his neighbor to the right. Each cryptographer then observes the two coins that he can see, and announces *agree* or *disagree*. If a cryptographer is not paying, he will announce *agree* if the two sides are the same and *disagree* if they are not. However, if he is paying then he will say the opposite. It can be proved that if the number of *disagrees* is even, then the master is paying; otherwise, one of the cryptographers is paying. Furthermore, if one of the cryptographers is

paying, then neither an external observer nor the other two cryptographers can identify, from their individual information, who exactly is paying, assuming that the coins are fair.

The anonymity of the protocol is based on the fact that for each announcement of the cryptographers under a payer i, there is a different configuration of the coins producing the same announcement under a different payer j. Moreover, if the coins are fair, these configurations have the same probability. However, if the coins are not fair then strong anonymity is lost, since the coin configurations are not equally probable and the attacker can assign higher probability to a particular cryptographer. In the extreme case of totally biased coins, we can expose the payer by comparing the announcements to the coin values (which are fixed). In the following section we measure the degree of anonymity of the protocol as a function of the bias of the coins.

6.1.2 Model-checking the protocol

To measure the degree of anonymity of a system, we start by identifying the set of anonymous events, which depend on what the system is trying to hide. In protocols where one user performs an action of interest (like paying in our example) and we want to protect his identity, the set $\mathcal{A}$ would be the same as the set I of the users of the protocol. In the dining cryptographers, we take $\mathcal{A} = \{c_1, c_2, c_3, m\}$ where c_i means that cryptographer i is paying and m that the master is paying. In protocols where k users can perform the action of interest simultaneously at each protocol execution, $\mathcal{A}$ would contain all k-tuples of elements of I. Another interesting case are MIX protocols, in which we are not interested in protecting the fact that someone sent a message (this is indeed detectable), but instead, the link between the sender and the receiver, when k senders send messages to k receivers simultaneously. In that case we consider the sets I_s, I_r of senders and receivers respectively, and take $\mathcal{A}$ to contain all k-tuples of pairs (a, a') where $a \in I_s, a' \in I_r$.

Then the set of observable events should also be defined, based on the visible actions of the protocol and on the various assumptions made about the attacker. In the dining cryptographers, we consider for simplicity the case where all the cryptographers are honest and the attacker is an external observer (the case of corrupted cryptographers can be treated similarly). Since the coins are only visible to the cryptographers, the only observables of the protocol are the announcements of *agree*/*disagree*. So the set of observable events will contain all possible combinations of announcements, that is $\mathcal{O} = \{aaa, aad, \ldots, ddd\}$ where a means *agree* and d means *disagree*.

If some information about the anonymous events is revealed intentionally then we should consider using relative anonymity (see Section 3.1). In the dining

	daa	*ada*	*aad*	*ddd*	*aaa*	*dda*	*dad*	*add*
c_1	0.25	0.25	0.25	0.25	0	0	0	0
c_2	0.25	0.25	0.25	0.25	0	0	0	0
c_3	0.25	0.25	0.25	0.25	0	0	0	0
m	0	0	0	0	0.25	0.25	0.25	0.25

	daa	*ada*	*aad*	*ddd*	*aaa*	*dda*	*dad*	*add*
c_1	0.37	0.21	0.21	0.21	0	0	0	0
c_2	0.21	0.37	0.21	0.21	0	0	0	0
c_3	0.21	0.21	0.37	0.21	0	0	0	0
m	0	0	0	0	0.37	0.21	0.21	0.21

Fig. 4. The channel matrices for probability of heads $p = 0.5$ (top) and $p = 0.7$ (bottom)

cryptographers, the information about whether the payer is a cryptographer or not is revealed by design (this is the purpose of the protocol). If, for example, the attacker observes *aaa* then he concludes that the anonymous event that happened is m since the number of *disagree* is even. To model this fact we use the conditional capacity and we take $\mathcal{R} = \{m, c\}$ where m means that the master is paying and c that one of the cryptographers is paying.

After defining $\mathcal{A}, \mathcal{O}, \mathcal{R}$ we should model the protocol in some formal probabilistic language. In our example, we modeled the dining cryptographers in the language of the PRISM model-checker ([30]), which is essentially a formalism to describe Markov Decision Processes. Then the channel matrix of conditional probabilities $p(o|a)$ must be computed, either by hand or using an automated tool like PRISM. In the case of relative anonymity, the probabilities $p(o|a)$ and $p(r|a, o)$ are needed for all a, o, r. However, in our example, R is a deterministic function of both A and O, so by Theorem 3 we can compute the conditional capacity as the maximum capacity of the sub-channels for each value of R individually. For $R = m$ the sub-channel has only one input value, hence its capacity is 0. Therefore the only interesting case is when $R = c$. In our experiments, we use PRISM to compute the channel matrix, while varying the probability p of each coin yielding heads. PRISM can compute the probability of reaching a specific state starting from a given one. Thus, each conditional probability $p(o|a)$ is computed as the probability of reaching a state where the cryptographers have announced o, starting from the state where a is chosen. In Fig. 4 the channel matrix is displayed for $p = 0.5$ and $p = 0.7$.

Finally, from the matrix, the capacity can be computed in two different ways.

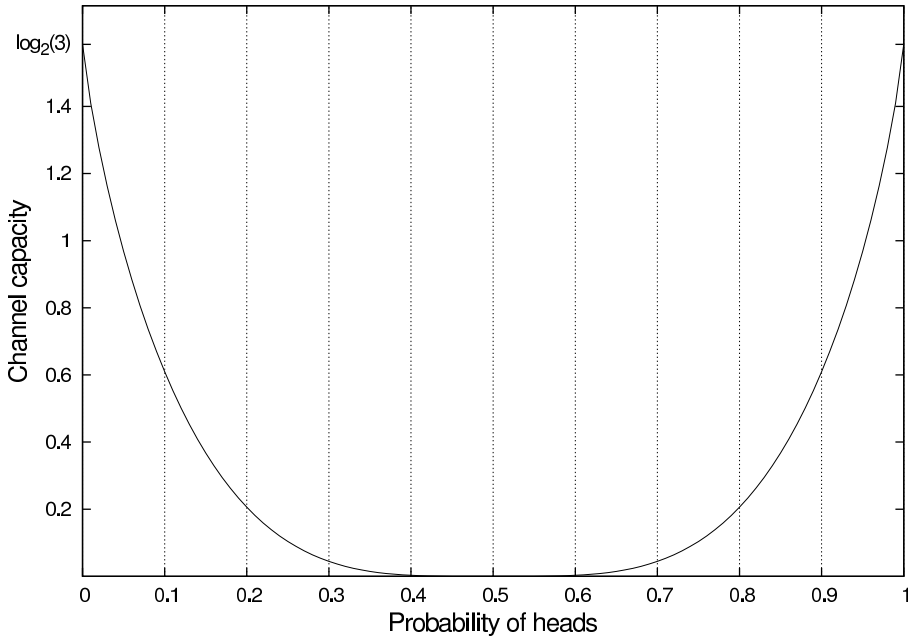


Fig. 5. The degree of anonymity in the Dining Cryptographers as a function of the coins' probability to yield heads.

Either by using the general Arimoto-Blahut algorithm, or by using Theorem 7 which can be applied because the matrix is partially symmetric. The resulting graph is displayed in Fig. 5. As expected, when $p = 0.5$ the protocol is strongly anonymous and the relative loss of anonymity is 0. When p approaches 0 or 1, the attacker can deduce the identity of the payer with increasingly high probability, so the capacity increases. In the extreme case where the coins are totally biased the attacker can be sure about the payer, and the capacity takes its maximum value of $\log 3$.

6.2 Crowds

6.2.1 Description of the protocol

This protocol allows Internet users to perform web transactions without revealing their identity. When a user communicates with a web server to request a page, the server can know from which IP address the request was initiated. The idea, to obtain anonymity, is to randomly route the request through a crowd of users. The routing protocol ensures that, even when a user appears to send a message, there is a substantial probability that he is simply forwarding it for somebody else.

More specifically a crowd is a group of m users who participate in the protocol. Some of the users may be corrupted which means they can collaborate in order to reveal the identity of the originator. Let c be the number of such users and

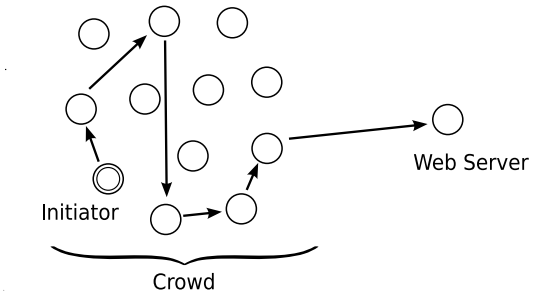


Fig. 6. The Crowds protocol

p_f a parameter of the protocol, explained below. When a user, called the *initiator* or *originator*, wants to request a web page he must create a *path* between him and the server. This is achieved by the following process, also displayed in Figure 6.

- The initiator selects randomly a member of the crowd (possibly himself) and forwards the request to him. We will refer to this latter user as the *forwarder*.
- A forwarder, upon receiving a request, flips a biased coin. With probability $1 - p_f$ he delivers the request directly to the server. With probability p_f he selects randomly, with uniform probability, a new forwarder (possibly himself) and forwards the request to him. The new forwarder repeats the same procedure.

The response from the server follows the same route in the opposite direction to return to the initiator. Each user is considered to have only access to the traffic routed through him, so he cannot intercept messages addressed to other users.

It is easy to see that, with respect to the web server, the protocol offers strong anonymity. The more interesting case, however, is the anonymity wrt a corrupted user that participates in the protocol. In this case, the initiator might try to forward the message to the attacker, so the latter can gain more information than the end server. We say that a user is *detected* if he sends a message to a corrupted user. Then it is clear that the initiator, since he always appears in a path, is more likely to be detected than the rest of the users. Thus detecting a user increases his probability of being the initiator, so strong anonymity cannot hold. However, if the number of corrupted users is not too big, the protocol can still satisfy probable innocence, meaning that the detected user is still less likely to be the originator than all the other users together, even though he is more likely than each other user individually.

Consider a Crowds instance of m users of which n are honest and $c = m - n$ are corrupted. Since anonymity makes sense only for honest users we define $\mathcal{A} = \{a_1, \ldots, a_n\}$ where a_i means that user i is the initiator of the message. The set of observables $\mathcal{O}$ depends on the attacker model, we could measure sender anonymity wrt the end server or wrt the corrupted users of the protocol, here we only consider the latter which is more interesting. The only thing that a corrupted user can observe is a request to forward a message, coming from another user of the protocol. Moreover, as it is usually the case in the analysis of Crowds ([31,32]), we assume that a corrupted user will never forward a message sent to him since by doing so he cannot learn more information about the actual initiator. Thus, there is at most one observed user (the one who sent the message to the corrupted user) and it is always an honest one. So we define $\mathcal{O} = \{o_1, \ldots, o_n\}$ where o_i means that the user i forwarded a message to a corrupted user.

Again, the channel matrix $p(o|a)$ can be computed either analytically or by means of a model-checking tool like PRISM. The advantage of the second approach is that with minimal changes we could compute the matrix for any network topology, not only for the usual clique network, which is much more difficult to do analytically. In fact, in [33] we use PRISM to compute the matrix of Crowds in a grid network. Since PRISM can only check finite-state models, we need to model Crowds as a finite-state system, even though its executions are infinite. We use a similar model as in [31] where a state is defined by the user who currently possesses the message, independently form the path that the message followed to arrive there, so the number of states is finite. In order for $p(\cdot|a)$ to be a distribution over $\mathcal{A}$, we normalize all elements by dividing with the total probability of observing any user. This corresponds to computing all probabilities conditioned on the event that some user has been observed, which is reasonable since if no user is observed at all then anonymity is not an issue.

From the matrix we can compute the capacity, for the case of a clique network, using Theorem 5 since the matrix is symmetric. As a consequence we only need one row of the matrix, so we can only compute a single one to speed up model-checking. For non-clique networks we can still compute the capacity using the Arimoto-Blahut algorithm.

The resulting graph is displayed in Fig. 7. We have plotted the capacity of three Crowds instances while varying the probability p_f of forwarding a message in the protocol. All instances have 50 honest users while the number of corrupted ones is 10, 20 and 30 respectively. Firstly, we see that the whole graph of the capacity is smaller when the number of corrupted users is smaller, which is ex-

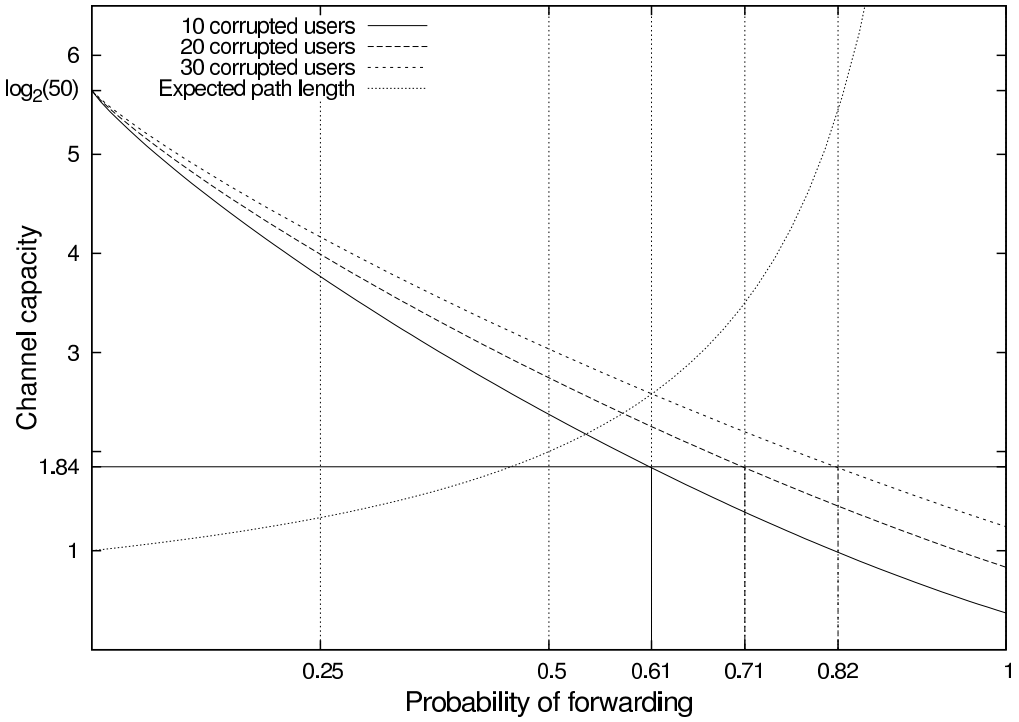


Fig. 7. The degree of anonymity for Crowds as a function of the probability p_f of forwarding a message. Three instances are displayed, with 50 honest users and 10, 20 and 30 corrupted ones. The expected path length is also displayed as a function of p_f.

pected since more corrupted users means higher probability of getting detected in the first round. When $p_f = 0$ then all instances have maximum capacity $\log_2 50$, meaning no anonymity at all, since, if forwarding never happens then the detected user is always the initiator.

For each instance we also indicate the minimum value of p_f required to satisfy probable innocence, given by the equation $m = \frac{p_f}{p_f - \frac{1}{2}}(c+1)$ ([10]). This value is different for each instance (since m, c are different) however at this value all instances have the same capacity $C = H(p_u) - H(p_{1/2}) \approx 1.8365$ where p_u is a uniform distribution over $\mathcal{A}$ and $p_{1/2}$ is a distribution that assigns probability $1/2$ to one user, and uniform to all the others.

Finally, the expected length of the path to the server, equal to $\frac{1}{1-p_f}$ (as shown in [10]) is displayed. As we can see from the graph there is a trade-off between performance (expected path length) and anonymity (capacity) when selecting a value for p_f. Given the maximum number of corrupted users that we want to consider, we can use the graph to find a value for p_f that offers acceptable capacity with a reasonable expected path length. The quantitative aspect of the capacity is important in this case, since it provides more detail about the connection between the degree of anonymity and p_f, even in areas where

probable innocence is always satisfied or violated.

In these two examples, we see how the various results of this paper fit together when we analyze an anonymity protocol. We model the protocol by considering the anonymous events $\mathcal{A}$, the observable events $\mathcal{O}$, and the matrix $p(o|a)$. In this framework, the loss of anonymity (Definition 1) gives an intuitive measure of the anonymity degree of the protocol. In the case of relative anonymity the revealed information $\mathcal{R}$ and the probabilities $p(r|a, o)$ need also to be considered, and the relative loss of anonymity (Definition 2) needs to be computed. Theorem 3 greatly reduces the size of the problem since we only need to compute the traditional capacity of the sub-matrices of $p(o|a)$. Computing the capacity is further simplified by partial symmetry, we only need to compute one row of the matrix and the computation of the capacity is a very simple operation on this row. Finally, the actual computation of the conditional probabilities can be fully automated using a model-checking tool like PRISM.

References

[1] S. Schneider, A. Sidiropoulos, CSP and anonymity, in: Proc. of the European Symposium on Research in Computer Security (ESORICS), Vol. 1146 of Lecture Notes in Computer Science, Springer, 1996, pp. 198–218.

[2] P. Y. Ryan, S. Schneider, Modelling and Analysis of Security Protocols, Addison-Wesley, 2001.

[3] P. F. Syverson, S. G. Stubblebine, Group principals and the formalization of anonymity, in: World Congress on Formal Methods (1), 1999, pp. 814–833.

[4] J. Y. Halpern, K. R. O'Neill, Anonymity and information hiding in multiagent systems, Journal of Computer Security 13 (3) (2005) 483–512.

[5] D. Hughes, V. Shmatikov, Information hiding, anonymity and privacy: a modular approach, Journal of Computer Security 12 (1) (2004) 3–36.

[6] L. Sweene, k-anonymity: A model for protecting privacy, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (5) (2002) 557–570.

[7] L. Sweene, Achieving k-anonymity privacy protection using generalization and suppression, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (5) (2002) 571–588.

[8] D. Chaum, The dining cryptographers problem: Unconditional sender and recipient untraceability, Journal of Cryptology 1 (1988) 65–75.

[9] M. Bhargava, C. Palamidessi, Probabilistic anonymity, in: M. Abadi, L. de Alfaro (Eds.), Proceedings of CONCUR, Vol. 3653 of Lecture Notes in Computer Science, Springer, 2005, pp. 171–185.

[10] M. K. Reiter, A. D. Rubin, Crowds: anonymity for Web transactions, ACM Transactions on Information and System Security 1 (1) (1998) 66–92.

[11] K. Chatzikokolakis, C. Palamidessi, Probable innocence revisited, Theoretical Computer Science 367 (1-2) (2006) 123–138.

[12] A. Serjantov, G. Danezis, Towards an information theoretic metric for anonymity., in: R. Dingledine, P. F. Syverson (Eds.), Proceedings of the workshop on Privacy Enhancing Technologies (PET) 2002, Vol. 2482 of Lecture Notes in Computer Science, Springer, 2002, pp. 41–53.

[13] C. Díaz, S. Seys, J. Claessens, B. Preneel, Towards measuring anonymity, in: R. Dingledine, P. F. Syverson (Eds.), Proceedings of the workshop on Privacy Enhancing Technologies (PET) 2002, Vol. 2482 of Lecture Notes in Computer Science, Springer, 2002, pp. 54–68.

[14] I. S. Moskowitz, R. E. Newman, D. P. Crepeau, A. R. Miller, Covert channels and anonymizing networks., in: S. Jajodia, P. Samarati, P. F. Syverson (Eds.), WPES, ACM, 2003, pp. 79–88.

[15] I. S. Moskowitz, R. E. Newman, P. F. Syverson, Quasi-anonymous channels, in: IASTED CNIS, 2003, pp. 126–131.

[16] Y. Zhu, R. Bettati, Anonymity vs. information leakage in anonymity systems, in: Proc. of ICDCS, IEEE Computer Society, 2005, pp. 514–524.

[17] M. R. Clarkson, A. C. Myers, F. B. Schneider, Belief in information flow, Journal of Computer SecurityTo appear. Available as Cornell Computer Science Department Technical Report TR 2007-207.

[18] Y. Deng, J. Pang, P. Wu, Measuring anonymity with relative entropy, in: T. Dimitrakos, F. Martinelli, P. Y. A. Ryan, S. A. Schneider (Eds.), Proceeddings of the of the 4th International Workshop on Formal Aspects in Security and Trust, Vol. 4691 of Lecture Notes in Computer Science, Springer, 2006, pp. 65–79.

[19] J. McLean, Security models and information flow, in: IEEE Symposium on Security and Privacy, 1990, pp. 180–189.

[20] 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), IEEE, Washington - Brussels - Tokyo, 1991, pp. 21–35.

[21] D. Clark, S. Hunt, P. Malacaria, Quantitative analysis of the leakage of confidential data, in: Proc. of QAPL 2001, Vol. 59 (3) of Electr. Notes Theor. Comput. Sci, Elsevier Science B.V., 2001, pp. 238–251.

[22] D. Clark, S. Hunt, P. Malacaria, Quantified interference for a while language, in: Proc. of QAPL 2004, Vol. 112 of Electr. Notes Theor. Comput. Sci, Elsevier Science B.V., 2005, pp. 149–166.

[23] G. Lowe, Quantifying information flow, in: Proc. of CSFW 2002, IEEE Computer Society Press, 2002, pp. 18–31.

[24] U. M. Maurer, Authentication theory and hypothesis testing, IEEE Transactions on Information Theory 46 (4) (2000) 1350–1356.

[25] A. D. Pierro, C. Hankin, H. Wiklicky, Approximate non-interference, Journal of Computer Security 12 (1) (2004) 37–82.

[26] A. D. Pierro, C. Hankin, H. Wiklicky, Measuring the confinement of probabilistic systems, Theoretical Computer Science 340 (1) (2005) 3–56.

[27] T. M. Cover, J. A. Thomas, Elements of Information Theory, John Wiley & Sons, Inc., 1991.

[28] A. Sabelfeld, D. Sands, Probabilistic noninterference for multi-threaded programs, in: Proc. of CSFW 2000, IEEE Computer Society Press, 2000, pp. 200–214.

[29] Y. Deng, C. Palamidessi, J. Pang, Weak probabilistic anonymity, in: Proceedings of the 3rd International Workshop on Security Issues in Concurrency (SecCo), Vol. 180 (1) of Electronic Notes in Theoretical Computer Science, Elsevier Science B.V., 2007, pp. 55–76.

[30] M. Z. Kwiatkowska, G. Norman, D. Parker, PRISM 2.0: A tool for probabilistic model checking, in: Proceedings of the First International Conference on Quantitative Evaluation of Systems (QEST) 2004, IEEE Computer Society, 2004, pp. 322–323.

[31] V. Shmatikov, Probabilistic analysis of anonymity, in: 15th IEEE Computer Security Foundations Workshop (CSFW), 2002, pp. 119–128.

[32] M. Wright, M. Adler, B. Levine, C. Shields, An analysis of the degradation of anonymous protocols, in: ISOC Network and Distributed System Security Symposium (NDSS), 2002.

[33] K. Chatzikokolakis, C. Palamidessi, P. Panangaden, Probability of error in information-hiding protocols, in: Proceedings of the 20th IEEE Computer Security Foundations Symposium (CSF20), IEEE Computer Society, 2007, pp. 341–354.