Anonymity Protocols as Noisy Channels *

Konstantinos Chatzikokolakis Catuscia Palamidessi
INRIA and LIX, École Polytechnique, Palaiseau, France
{kostas,catuscia}@lix.polytechnique.fr

Prakash Panangaden
School of Computer Science, McGill University, Montreal, Quebec, Canada
prakash@cs.mcgill.ca

Abstract. We propose a framework in which anonymity protocols are interpreted as particular kinds of channels, and the degree of anonymity provided by the protocol as the converse of the channel's capacity. We justify this view in terms of the Bayesian probability of error that limits the adversary's capability of testing the protocol to infer the user's identity. 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.

1 Introduction

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

* 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.

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 propose a definition of the degree of anonymity of a protocol in terms of the information-theoretic notion of *capacity* of the protocol, seen as channel. We also define a more general notion, that we call *relative capacity*, which naturally models the case in which some loss of an anonymity is allowed by design.

We justify our proposal by showing the relation with the knowledge that an attacker can gain on the anonymous actions (the channel's inputs), from the observables (the channel's outputs) and the matrix of the conditional probabilities associated to the channel. In particular, we consider attackers following the

so-called Bayesian approach, and we show bounds on the Bayesian probability of error regarding the probabilistic information that the attacker can acquire.

We then compare our proposal with various probabilistic notions of anonymity given in the past, in particular perfect anonymity, group anonymity, and probable innocence. Finally, we show that the condition of probable innocence corresponds to a certain information-theoretic bound.

1.2 Related work

Probabilistic definitions of anonymity have been explored in [3, 11, 1, 16, 2]. 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 [18, 9]. 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, and the purpose is to measure the amount of information that can be transfered through these channels. In this paper, we consider the channel to be the protocol itself, as an abstraction that allows us to measure anonymity.

Another approach close in spirit to ours is the one of [8]. In this work, the authors use the notion of relative entropy 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 [8] 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 [13, 10, 4, 5, 12] in which the 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 our findings are applicable more or less directly also to the field of non-interference.

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 relative 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. Finally, in Section 5, we relate our framework to other probabilistic approaches to anonymity.

The proofs of all the results (except those that trivially follow from known results in literature) are in the appendix.

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 [6] 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. It 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 join probability $p(a, o)$ of the particular values taken by A, O 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

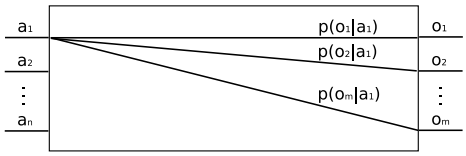

Fig. 1. An anonymity channel

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 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 of 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 define 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.

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 [3, 11, 1]. However, as we discuss in section 5.3, our framework also captures weaker notions of anonymity.

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

Fig. 2. A simple elections protocol

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 MIXing technique. 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 kind 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.

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 *relative loss of anonymity.*

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.

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 partitions a random variable 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 1. *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 partitions both A and O 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}.

3.2 Computing the channel's capacity

In general, there is no formula to compute the capacity of an arbitrary channel. 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 3. *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 2 ([6], 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 2 holds for both classes.

In anonymity protocols, we expect all rows of the protocol's matrix to be permutations of each other since all users are executing the same protocol. On the other hand, the columns are not necessarily permutations of each other. Some symmetry is expected: if an observable o_1 is produced with probability p under user a_1, it is reasonable to assume that under a_2 there will be some other observable o_2 produced with the same probability. However, we can have observables that are produced with equal probability by all users. Clearly, these "constant" columns cannot be the permutation of a non-constant one 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 4. *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 2 to the case of partial symmetry.

Theorem 3. *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 $\mathbf{r}_s$.

Note that Theorem 3 is a generalization of Theorem 2. 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 2 from Theorem 3.

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 corresponds to allowing the attacker to force the user to redo the action. For instance, in Crowds, the attacker can intercept the message and destroy it, thus obliging the sender to resend it. 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 attacker forces the users to execute the protocol n times with the same a as input event, and tries to infer a from the n observable outputs of the protocol executions. Let O_1, O_2, ..., O_n represent the random variables corresponding to the observations made by the attacker, and let $\boldsymbol{o}$ denote a sequence of observed outputs $o_1, o_2, \ldots o_n$. As stated above, we assume that O_1, O_2, ..., 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

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

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 $E_n : \mathcal{A} \to \mathcal{O}^n$ be the function that gives the *error region* of f_n when $a \in \mathcal{A}$ has occurred, namely:

$$E_f(a) = \{\boldsymbol{o} \in \mathcal{O}^n \mid f(\boldsymbol{o}) \neq a\}$$

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 when $a \in \mathcal{A}$ has occurred, namely:

$$\eta(a) = \sum_{\boldsymbol{o} \in E_f(a)} p(\boldsymbol{o}|a)$$

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 [6], Section 12.8).

Definition 5. *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* Bayesian probability of error P_{f_n} *is defined as the probability weighted sum over $\mathcal{A}$ of the individual probabilities of error. Namely:*

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

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 $\boldsymbol{o} \in \mathcal{O}^n$, tries to maximize the a posteriori probability of the hypothesis a w.r.t. $\boldsymbol{o}$. The a posteriori probability of a w.r.t. $\boldsymbol{o}$ is given by Bayes theorem (aka Bayes Inversion Rule):

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

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

Definition 6. *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 $\boldsymbol{o} \in \mathcal{O}^n$, $f_n(\boldsymbol{o}) = a$ implies $p(\boldsymbol{o}|a)p(a) \geq p(\boldsymbol{o}|a')p(a')$ 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 1. *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 $P_{f_n} \leq P_{h_n}$.*

4.1 Independence from the input distribution

The definition of the Bayesian decision functions depends on the a priori probability distribution of 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 Bayesian probability of error do not depend on the distribution of A.

The following definition establishes the condition on the matrix.

Definition 7. *Given an anonymity protocol $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$, we say that such protocol is Bayesian-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 $a \neq a'$.*

We will now show that if a protocol is Bayesian-determinate, then in the definition of the decision functions the distribution on A eventually washes out. The intuition is that, in the comparison between $p(\boldsymbol{o}|a)p(a)$ and $p(\boldsymbol{o}|a')p(a')$, the factor $p(a)p(a')$ is dominated by the factor $p(\boldsymbol{o}|a)p(\boldsymbol{o}|a')$, for n large enough, provided that the latter is different from 1.

Proposition 2. *Given a Bayesian-determinate anonymity protocol $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$, for any distribution $p(\cdot)$ on A, and for any $\epsilon > 0$, there exists n such that for each Bayesian decision functions f_n there exists a decision function $g_n : \mathcal{O}^n \to \mathcal{A}$ such that $g_n(\boldsymbol{o}) = a$ implies $p(\boldsymbol{o}|a) \geq p(\boldsymbol{o}|a')$ for all $a' \in \mathcal{A}$, and such that g_n approximates f_n, in the sense that the probability of the set $\{\boldsymbol{o} \in \mathcal{O}^n \mid f_n(\boldsymbol{o}) \neq g_n(\boldsymbol{o})\}$ is smaller than ϵ.*

Proposition 2 allows us to define a decision function, for n sufficiently large, by comparing only the probabilities $p(\boldsymbol{o}|a)$ 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.

4.2 Bounds on the Bayesian 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 Bayesian-determinate:

Proposition 3. *Given a Bayesian-determinate anonymity protocol $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$, for any distribution $p(\cdot)$ on A, and for any $\epsilon, \epsilon' > 0$, there exists n such that the property*

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

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

Proposition 3 and its proof tell us that, in case of Bayesian-determinate matrices, there is essentially only one decision function, and it is value is determined, for n sufficiently large, by the a for which $p(\boldsymbol{o}|a)$ is greatest.

Consider now the converse 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 $\boldsymbol{o} \in \mathcal{O}^n$ such that $n(o, \boldsymbol{o}) = p(a|a) = p(o|a')$ for all $o \in \mathcal{O}$, the value of g_n is not uniquely determined, because $p(\boldsymbol{o}|a_1)$ and $p(\boldsymbol{o}|a_2)$ are both maximals. Assuming that we chose arbitrarily between a_1 and a_2, and that the probability of choosing the wrong one is uniformly distributed, we have that the Bayesian

probability of error is bound from below as follows: $P_{g_n} = \sum_{a \in \mathcal{A}} p(a)\eta(a) \geq p(a_1)1/2 + p(a_2)1/2$.

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

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 $p(a)p(\boldsymbol{o}|a)$ (for the $\boldsymbol{o}$ s.t. $n(o, \boldsymbol{o}) = p(a|a) = p(o|a')$ for all $o \in \mathcal{O}$) 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 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 on A. From information theory we know that this is the case iff A and O are independent (cfr. [6], page 27). Hence we have the following characterization:

Proposition 4. *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 [1] 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 [3] and it is called *conditional anonymity* in [11].

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 $p(\boldsymbol{o}|a_1) = p(\boldsymbol{o}|a_2)$ for all $a_1, a_2 \in \mathcal{A}$, and $\boldsymbol{o} \in \mathcal{O}^n$. Consequently the observations are of no use for the attacker to infer the anonymous event, i.e. to define the "right" $g_n(\boldsymbol{o})$, since all $p(\boldsymbol{o}|a)$ are maximal. Assuming a uniform distribution in assigning a value to $g_n(\boldsymbol{o})$, the Bayesian 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 [3], under the assumption that it is always one of the cryptographers who pays, and that the coins are fair.

5.2 Relative capacity 0: strong group anonymity

Group anonymity usually indicates the situation in which the users are divided in groups, and the protocol allows to figure out the group which the culprit belongs to, although it tries to conceal which user in the group is the culprit.

Such situation corresponds to having a partition on $\mathcal{A}$ and $\mathcal{O}$, see Section 3.1. The case of relative 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 $\boldsymbol{o} \in \mathcal{O}^n$, there exists exactly one group r_i of a's such that $p(\boldsymbol{o}|a) > 0$, and $p(\boldsymbol{o}|a_1) = p(\boldsymbol{o}|a_2)$ for all a_1, a_2 in r_i. Hence the attacker knows that the "right" value of $g_n(\boldsymbol{o})$ is an a in r_i, but he does not know exactly which. 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 Bayesian probability of error is $(|\mathcal{A}_r| - 1)/|\mathcal{A}_r|$ where $r \in \mathcal{R}$ determines the set of maximal cardinality in $\mathcal{A}$.

An example of protocol with relative capacity 0 is the dining cryptographers in a generic graph [3], under the assumption that the coins are fair. The groups correspond to the connected components of the graph.

The notion of strong group anonymity seems also related to the notion of equivalence classes in [17]. 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 [16] for Crowds, a system based on communicating a message from the originator to the receiver through a sequence of users acting as forwarders. 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 [16, 11, 2] that try to formally express this notion, see [2] for details. In this section we discuss the relation between these definitions and the channel capacity.

Definition of Chatzikokolakis and Palamidessi The definition of [2] tries to combine the other two by considering both the probability of producing some observable and the attackers confidence after the observation. This definition considers the probability of two anonymous evens 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 if

$$(n-1) \geq \frac{p(o|a)}{p(o|a')} \quad \forall o \in \mathcal{O}, \forall a, a' \in \mathcal{A} \tag{2}$$

where $n = |\mathcal{A}|$. In [2] 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 8. *An anonymity protocol* $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$ *satisfies* partial anonymity *if there is a constant* γ *such that*

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

A similar notion is called *weak probabilistic anonymity* in [7].

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 4. *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*

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

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 4 about strong probabilistic anonymity ($\gamma = 1$) from Theorem 4. A bound for probable innocence can be obtained by taking $\gamma = n-1$, so Theorem 4 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 guarrantee. In fact, condition 2 does not prevent the function $p(\boldsymbol{o}|\cdot)$ from having a maximum with probability close to 1, for a sufficiently long sequence of observables $\boldsymbol{o}$. So we can define $g_n(\boldsymbol{o})$ to be such maximum, and we have that the Bayesian error corresponding to g_n goes to 0. The only exception is when two (or more) raws a_1, a_2 are equal and correspond to maximals. Imposing this condition for all anonymous actions is equivalent to require strong anonymity. In conclusion, possible innocence maintains an upper bound on anonymity through protocol repetition only if the system is strongly anonymous. This result generalizes the one expressed by Proposition 17 in [2]: In the latter, the same conclusion is drawn, but the tests are limited to the observable sequences of the form $o, o, \ldots, o$.

Definition of Reiter and Rubin In [16] Reiter and Rubin give a formalization of probable innocence for the Crowds protocol, which limits the probability of

	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

detection, that is the probability of a certain observable that reveals each sender. The definition requires the probability of these observables to be less that one half. A protocol satisfies RR-probable innocence if

$$p(o|a) \leq \frac{1}{2} \quad \forall o \in \mathcal{O}, \forall a \in \mathcal{A} \tag{3}$$

In [2] it is argued that this definition 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 there are some special symmetries under which RR-probable innocence is equivalent to CP-probable innocence so a bound on the capacity can be obtained.

Definition of Halpern and O'Neill In [11] 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 event should be at most one half, under any observation. A protocol satisfies HO-probable innocence if

$$p(a|o) \leq \frac{1}{2} \quad \forall o \in \mathcal{O}, \forall a \in \mathcal{A} \tag{4}$$

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 [2], 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 5. *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*

$$I(A;O) \leq H(A) - 1$$

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

References

1. Bhargava, M. and Palamidessi, C. Probabilistic anonymity. In *Proc. of CONCUR*, *LNCS* 3653, pp. 171–185. Springer, 2005. www.lix.polytechnique.fr/∼catuscia/papers/Anonymity/concur.pdf.
2. Chatzikokolakis, K. and Palamidessi, C. Probable innocence revisited. *Theor. Comp. Sci.*, 2006. To appear. www.lix.polytechnique.fr/∼catuscia/papers/Anonymity/reportPI.pdf.
3. Chaum, D. The dining cryptographers problem: Unconditional sender and recipient untraceability. *Journal of Cryptology*, 1:65–75, 1988.
4. Clark, D., Hunt, S., and Malacaria, P. Quantitative analysis of the leakage of confidential data. In *Proc. of QAPL 2001*, *ENTCS* 59 (3). Elsevier, 2001.
5. Clark, D., Hunt, S., and Malacaria, P. Quantified interference for a while language. In *Proc. of QAPL 2004*, *ENTCS* 112, pp. 149–166. Elsevier, 2005.
6. Cover, T.M. and Thomas, J.A. *Elements of Information Theory*. Wiley, 1991.
7. Deng, Y., Palamidessi, C., and Pang, J. Weak probabilistic anonymity. In *Proc. of SecCo 2005*, *ENTCS*. Elsevier, 2005. www.lix.polytechnique.fr/∼catuscia/papers/Anonymity/report_wa.pdf.
8. Deng, Y., Pang, J., and Wu, P. Personal Communication, 2006
9. Díaz, C., Seys, S., Claessens, J., and Preneel, B. Towards measuring anonymity. In *Proc. of PET 2002*, *LNCS* 2482, pp. 54–68. Springer, 2002.
10. Gray, J.W. III. Toward a mathematical foundation for information flow security. In *Proc. of SSP'91*, pp. 21–35, IEEE, 1991.
11. Halpern, J.Y. and O'Neill, K.R. Anonymity and information hiding in multiagent systems. *Journal of Computer Security*, 13(3):483–512, 2005.
12. Lowe, G. Quantifying information flow. In *Proc. of CSFW 2002*, pages 18–31. IEEE, 2002.
13. McLean, J. Security models and information flow. In *Proc. of SSP'90*, pp. 180–189, IEEE, 1990.
14. Moskowitz, I.S., Newman, R.E., Crepeau, D.P., and Miller, A.R. Covert channels and anonymizing networks. In *Proc. of WPES'03*, pp. 79–88. ACM, 2003.
15. Moskowitz, I.S., Newman, R.E, and Syverson, P.S. Quasi-anonymous channels. In *Proc. of CNIS*, pages 126–131, IASTED, 2003.
16. Reiter, M.K. and Rubin, A.D. Crowds: anonymity for Web transactions. *ACM Transactions on Information and System Security*, 1(1):66–92, 1998.
17. Sabelfeld, A. and Sands, D. Probabilistic noninterference for multi-threaded programs. In *Proc. of CSFW 2000*, pp. 200–214. IEEEs, 2000.
18. Serjantov, A. and Danezis, G. Towards an information theoretic metric for anonymity. In *Proc. of PET 2002*, *LNCS* 2482, pages 41–53. Springer, 2002.

A Proofs

We present here the proofs that were omitted from the paper.

Theorem 1 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 partitions both A and O 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 $r \neq r'$. Since R partitions A, O we have $p(r|a) = 1$ and $p(r|o) = 0$. The idea is that, since a implies r, o cannot intersect with a unless it also intersects with r, which is impossible. More formally:

$$0 = p(o, r) = p(o, r, a) + p(o, r, \neg a) \geq p(o, r, a) = p(o, a)$$

since $p(a, r) = p(a)$. Thus $p(o, a) = 0$ which is only possible if $p(o|a) = 0$.

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

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

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

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

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

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:

$$I(A;O|R) = -\sum_{r\in\mathcal{R}} p(r)\Big[\sum_{a\in\mathcal{A}_r}\frac{p(a)}{p(r)}\log\frac{p(a)}{p(r)} - \sum_{o\in\mathcal{O}_r}\frac{p(o)}{p(r)}\sum_{a\in\mathcal{A}_r} p(a|o)\log p(a|o)\Big] \tag{5}$$

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 (5) can be written:

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

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 $C|R \leq d$. □

Theorem 3 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

$$\begin{aligned}
I(A;O) &= H(O) - H(O|A) \\
&= -\sum_{o\in\mathcal{O}} p(o) \log p(o) - H(\mathbf{r}) \\
&= -\sum_{o\in\mathcal{O}_s} p(o) \log p(o) - \sum_{o\in\mathcal{O}_n} p(o|a') \log p(o|a') - H(\mathbf{r}) \\
&= -\sum_{o\in\mathcal{O}_s} p(o) \log p(o) - H(\mathbf{r}_s) \\
&\leq -\sum_{o\in\mathcal{O}_s} \frac{p_s}{|\mathcal{O}_s|} \log \frac{p_s}{|\mathcal{O}_s|} - H(\mathbf{r}_s) \qquad (6) \\
&= p_s \log \frac{|\mathcal{O}_s|}{p_s} - H(\mathbf{r}_s) \qquad (7)
\end{aligned}$$

We constructed inequality (6) 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 (7) is achieved and it is the actual capacity of the channel. □

Proposition 1 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 Bayesian probability of error. See, for instance, [6], Section 12. □

Proposition 2 Given a Bayesian-determinate anonymity protocol $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$, for any distribution $p(\cdot)$ on A, and for any $\epsilon > 0$, there exists n such that for each Bayesian decision functions f_n there exists a decision function $g_n : \mathcal{O}^n \to \mathcal{A}$ such that $g_n(\boldsymbol{o}) = a$ implies $p(\boldsymbol{o}|a) \geq p(\boldsymbol{o}|a')$ for all $a' \in \mathcal{A}$, and such that g_n approximates f_n, in the sense that the probability of the set $\{\boldsymbol{o} \in \mathcal{O}^n \mid f_n(\boldsymbol{o}) \neq g_n(\boldsymbol{o})\}$ is smaller than ϵ.

Proof. Given a fixed input a and the corresponding sequence of observable outcomes $\boldsymbol{o} \in \mathcal{O}^n$, and given a value $o \in \mathcal{O}$, let $n(o, \boldsymbol{o})$ denote the number of o that occur in $\boldsymbol{o}$. By the strong law of large numbers ([6]), for any $\delta > 0$ the probability of the set $\{\boldsymbol{o} \in \mathcal{O}^n \mid \forall o \in \mathcal{O}\ |p(a)n(o, \boldsymbol{o})/n - p(a)p(o|a)| < \delta\}$ goes to 0 as n goes to ∞. Consequently, the probability of the set $S = \{\boldsymbol{o} \in \mathcal{O}^n \mid \forall \text{a'}\ p(a)p(\boldsymbol{o}|a) > p(a')p(\boldsymbol{o}|a')\}$ goes to 1 as n goes to ∞. In fact $p(a)p(\boldsymbol{o}|a)/p(a')p(\boldsymbol{o}|a') > 1$ iff

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

and

$$\begin{aligned}
\frac{1}{n}\log\frac{p(a)p(\boldsymbol{o}|a)}{p(a')p(\boldsymbol{o}|a')} &= \frac{1}{n}\log\frac{p(a)}{p(a')} + \frac{1}{n}\log\frac{p(\boldsymbol{o}|a)}{p(\boldsymbol{o}|a')} \\
&\underset{n\to\infty}{\longrightarrow} \frac{1}{n}\log\frac{p(\boldsymbol{o}|a)}{p(\boldsymbol{o}|a')} && \text{(since } \frac{1}{n}\log\frac{p(a)}{p(a')} \underset{n\to\infty}{\longrightarrow} 0\text{)} \\
&= \frac{1}{n}\log\prod_{i=1}^{n}\frac{p(o_i|a)}{p(o_i|a')} && \text{(by (1))} \\
&= \frac{1}{n}\sum_{i=1}^{n}\log\frac{p(o_i|a)}{p(o_i|a')} \\
&= \frac{1}{n}\sum_{o\in\mathcal{O}} n(o, \boldsymbol{o})\log\frac{p(o|a)}{p(o|a')} && \text{(by definition of } n(o, \boldsymbol{o})\text{)} \\
&\underset{n\to\infty}{\longrightarrow} \sum_{o\in\mathcal{O}} p(o|a)\log\frac{p(o|a)}{p(o|a')} && \text{(by the strong law of large numbers)} \\
&= D(p(\cdot|a) \parallel p(\cdot|a')) \\
&> 0 && \text{(by Bayesian-determinacy)}
\end{aligned}$$

Given a Bayesian decision function f_n, consider now the set $S' = \{\boldsymbol{o} \in \mathcal{O}^n \mid f_n(\boldsymbol{o}) = a\}$. 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 the probability of the set $\{\boldsymbol{o} \in \mathcal{O}^n \mid g_n(\boldsymbol{o}) = a\}$ goes to 1 as n goes to ∞. We can therefore conclude that the same holds for the probability of the set $\{\boldsymbol{o} \in \mathcal{O}^n \mid g_n(\boldsymbol{o}) = f_n(\boldsymbol{o})\}$. □

Proposition 3 Given a Bayesian-determinate anonymity protocol $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$, for any distribution $p(\cdot)$ on A, and for any $\epsilon, \epsilon' > 0$, there exists n such that the property

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

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

Proof. Given $\boldsymbol{o} \in \mathcal{O}^n$, define $g_n(\boldsymbol{o}) = a$ iff a is the value of A for which $p(\boldsymbol{o}|a)$ is greatest. By following the same lines as in the proof of Proposition 2, we have that the set $\{\boldsymbol{o} \in \mathcal{O} \,|\, \forall a' \in \mathcal{A}\; p(\boldsymbol{o}|a) > p(\boldsymbol{o}|a')\}$ 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 $E_{g_n}\{\boldsymbol{o} \in \mathcal{O}^n \,|\, \exists a' \in \mathcal{A}\; p(\boldsymbol{o}|a) \leq p(\boldsymbol{o}|a')\}$ has probability smaller than ϵ'. Hence $\eta(a) < \epsilon'$ and $P_{g_n} = \sum_{a\in\mathcal{A}} p(a)\eta(a) < \epsilon'$. □

Theorem 4 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 2 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 $\forall a, a' \in \mathcal{A} : \gamma \geq \frac{p(o_1|a)}{p(o_1|a')}$, thus

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

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 $H(r)$ 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. Clearly, this operation lowers the entropy of the row since it increases the distance between the values (without any constraints the lowest entropy of 0 is given when one value of the row is 1 and all the others 0). 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

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

So $H(\mathbf{r})$ is a function $h(x)$ of only one variable x. We want to find the value x_0 which minimizes $h(x)$. 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

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

□

Proposition 5 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

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

□