A Combinatorial Approach for an Anonymity Metric

Dang Vinh Pham[1] and Dogan Kesdogan[1,2]

[1] Siegen University, Siegen, Germany
pham@fb5.uni-siegen.de
[2] NTNU-Norwegian University of Science and Technology, Trondheim, Norway
kesdogan@fb5.uni-siegen.de, dogan.kesdogan@q2s.ntnu.no

Abstract. A number of papers are suggested with the goal to measure the quality of anonymity of a given anonymity system. Most of them use the anonymity set as the basis for developing, reasoning about and applying measure. In this paper we argue that these approaches are premature. In this work we suggest to use the so called hypothesis set – a term derived from possibilistic information flow theory. Investigating the hypothesis set, it is possible to make the "protection structure" explicit and also define well known terms from measurement theory like scale and metric. We demonstrate our approach by evaluating the hypothesis set of the classical Chaumian Mix.

1 Introduction

One of the most important values in the information society is the information itself. Therefore, the protection of this precious good is a crucial task. A lot of research attention has been devoted to protecting the information contained *within* messages. This can be easily achieved today, by using encryption techniques. Here we focus not on the protection of such content data but rather on how to ensure the confidentiality of *traffic data*, i.e. information about communication relations. Protection of traffic data usually results in some form of anonymity. Such confidentiality is important, since third parties' unrestricted access to traffic data is considered an unacceptable invasion of both private and business lives. Several techniques are known to protect traffic data. However, there is still a lack of models to evaluate the level of protection these techniques can provide. There are two reasons to determine the level of protection. First, the quality of protection can be made visible to the user. Second, the mathematic model gives designers insight into the protection task.

In this paper we will study a specific system - the Mix system [1] that exposes the basis for an understanding of the abstract problem. Consequently, our focus is to investigate the abstract problem and the well known model of anonymity systems, the *anonymity set*, which can be used to model all traffic protection techniques [2]: Anonymity is the state of being not identifiable within a set of subjects, called the anonymity set.

All practical anonymity techniques leak some information about the users' communication peers in the anonymity set. As the number of observed anonymity sets increases, the uncertainty about the peer partners usually decreases, eventually reaching 0. At this point, there is enough information to determine the peer partners uniquely. According to this observation and inspired by the metric *Mean Time To Failure* (MTTF)

C. Boyd and J. González Nieto (Eds.): ACISP 2009, LNCS 5594, pp. 26–43, 2009.

[3], Shannon's *unicity distance* [4] and measurement theory [3], we define the metric *Mean Time To Deanonymization* (MTTD).

In [5] an attack algorithm called the *Hitting Set Attack* (HS-Attack) is proposed that can be used to measure MTTD for the MIX model. It was proven in [5] that HS-Attack requires the least number of observations that is necessary to uncover the peer partners of a user given a global passive attacker, who can observe any communication link from the sender to the Mix and from the Mix to the receivers. In this work we will investigate the HS-Attack for new structures to better understand how the attack evolves with increasing observations. We will mathematically approximate sharply this evolution from secure state into insecure state as a homomorphism of the "reality". Thereby we will provide mathematic answers to the following main questions:

- What is the average number of observations to disclose all of a user's peer partners?
- What is the average number of observations to disclose at least one of the user's peer partners?
- How likely is it to find a particular fraction of a user's peer partners in a random hypothesis after a given number of observations?

Question one was first considered with respect to inherent structures of the Mix- and attacker- model in [5]. They measured the mean number of observations to reach a necessary condition (called exclusivity) to disclose all of a user's peer partners. In contrast to [5], our approach is more comprehensive and granular, since it enables to directly model the number of observations to disclose any number of a user's peer partners.

In answering question two, we are the first to suggest the anonymity metric MTTD-1 that measures the time point, when the mix system's anonymity function leaks the first unambiguous information about a user's peer. This peer can be revealed by HS-Attack and we provide a mathematic measurement of the number of observations that the attacker needs to succeed. Each of a user's communication relationship is information theoretic anonymous, if it can be avoided that the attacker gains MTTD-1 observations. MTTD-1 extends the traditional measurement of anonymity that until now only considers the time point when all of a user's peer is disclosed.

Finally question three applies to the situation, when it is not possible to definitively identify any user's peer. Our mathematic model gives the probability that a random hypothesis contains a certain number of the user's contacts. The latter point shows that our approach opens the door to further analysis beyond the scope of unambiguous identification of peers and we are also the first to address this issue.

This paper is organised as follows. Section 2 will provide basic background information, related works and the used Mix meta model. Section 3 will describe the inherent structures and properties of hypotheses sets and we will contribute a precise mathematic model to describe them. Based on this model, we will derive analytical formulas to expose distinct anonymity states with respect to the Mix parameters and the observations of the attacker in Sect. 4. We will measure the mean number of observations to identify arbitrary subsets of a user's peer partners, the size of the hypothesis set and the probability to find a particular fraction of a user's partners in a randomly computed hypothesis. Section 5 will finally summarise our results and outline future works.

2 Background

Over the last years a handful of “anonymity metrics” have been proposed [6, 7, 8, 9, 10, 11, 12] that measure the information flow to the attacker. If information flows to the attacker, then it reduces the uncertainty of the attacker, which can be measured with some variant of Shannon’s entropy. All entropy measurements follow here a similar scheme: information flow occurs if the attacker is really uncertain[1] and afterwards fairly certain.

This kind of “macroscopic” measurement can easily be applied to different anonymity systems, since only the probability distributions have to be known. The paradigm “information flows when uncertainty decreases” is problematic as discussed in [13, 11]. Our approach is more microscopic. The attacker and his knowledge (his interpretation) is incorporated in the model. Thus, direct application of our approach to other anonymity techniques (e.g. Crowds [14]) is not given, i.e. it has to be adapted or rather redeveloped.

Again, we model the whole path from secure to insecure state where the uncertainty with each observation decreases (i.e. number of hypotheses decreases). Indeed, in each step we can measure the uncertainty by using entropy as suggested in the literature. This would be a concomitant measurement. However, we think that the other way around is not possible (without formal proof), since suggested entropy based approaches measure only the actual information flow but not how the security evolves with time. The reason for our approach is that we think that the goal of a security strength metric is *to quantify the attackers efforts* that are required to break a system’s security. Therefore, with respect to the general paradigm “the harder the successful attack the stronger the system” the actual entropy metric suggestions are premature[2].

Consequently, we define anonymity in terms of the attacker’s knowledge, using the well known trace-based approach [15]. An attacker observes and accumulates input and output events of the anonymity system. The measurement of anonymity depends on the knowledge (i.e. interpretation of the observations) of the attacker that we model with a set of hypotheses. The only assumption we make is that the user keeps the set of peer partners. Unlike the theoretic works in this area (see e.g. [16, 17, 18]), we are not interested to give a formal specification of the anonymity problem.

2.1 The Mix Model

We consider the Mix Model that was described in [5]. The *Mix* technique was proposed by David Chaum in 1981 [1]. Figure 1 shows the basic ingredients of this technique which consist of a set of senders S, a set of recipients R, and a Mix node. Note that S and R can be equal. All senders in S are connected to the Mix and the Mix itself is connected to all recipients in R by a communication network with reliable secure channels. A reliable secure channel does not

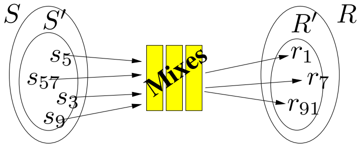

Fig. 1. Formal model

[1] E.g. a priori distribution is the uniform distribution.

[2] The draw back of our approach (as mentioned above) is the necessity of modelling explicitly the anonymity system and the attackers knowledge.

result in loss or duplication of transmitted messages, and guarantees authenticity, integrity, and confidentially of transmitted messages. The users and the Mix transmit messages by using the following protocols:

User Protocol: Users prepare their messages to be of constant length either by splitting long messages or by padding short messages to the specified length. Each message is encrypted twice with one time pads: first the message is encrypted using a shared secret between the sender and the intended recipient, and then it is encrypted using a shared secret between the sender and the Mix. The users send twice encrypted messages to the Mix.

Mix Protocol: A Mix collects b messages (called a *batch*) from distinct users, decrypts the messages, and outputs the decrypted messages in a batch in a different order than the order in that they were received (lexicographically sorted or randomly delayed). The output is broadcasted to all recipients. Furthermore, any incoming packet is compared with formerly received messages (i.e. by locally caching formerly received messages) in order to reject any duplicate messages.

The basic Mix technique described above can perfectly hide the communication relationships between senders and recipients of messages from everybody but the Mix and message senders. Even the act of sending or receiving can be perfectly hidden if the above protocol is applied in fixed time slots, and if every user supplies a fixed number of messages (perhaps some or all of them being dummy messages) to each slot and the whole output batch in a time slot is distributed to every user [19, 1, 20]. Pfitzmann [20] states that the Mix technique provides information-theoretic anonymity and unobservability based on complexity-theoretic secure cryptography.

The pure Mix technique. The "perfect" anonymity solution discussed above uses dummy messages and broadcasting. This solution is not followed widely in large networks such as the Internet, as justified in [5]. As a consequence, most current implementations and solutions use a variant of the perfect Mix solution by neither using dummy messages nor the broadcasting function. We refer to this kind of Mix techniques by the term *pure* Mix technique. Our pure Mix (also called threshold Mix) model is quite general and also Pool-Mixes can be mapped on it as shown in [10].

We consider a *global passive attacker* who is capable of observing all communication links simultaneously as described in [5]. This attacker model is also known as the Dolev-Yao model in the literature. Based on this attacker, we will use the following formal model of a pure Mix and information leakage therein for our analysis.

Formal Model of the Pure Mix Technique

- A communication system consists of a *set of senders* S, and a *set of recipients* R, and a Mix node (see Fig. 1). If a sender $s \in S$ communicates with a recipient $r \in R$, then we say that s and r are peer partners. If the roles of sender and receiver need to be distinguished, then we say that s is a peer sending partner of r and r is a peer recipient partner of s.

- In each *communication round*[3] a subset $S' \subseteq S$ of all senders S send a message to their peer partners. Let $R' \subseteq R$ be the set of intended recipients. The act of sending or receiving a message is not hidden among dummy messages.
- The size of the *sender anonymity set* is $|S'| = b$, where $1 < b \leqslant |S| = n$.
- The size of the *recipient anonymity set* is $|R'| \leqslant b$ since each sender sends exactly one message and several senders may communicate with the same recipient. The size of the recipient set is $|R| = N$.
- The information leakage X available to an attacker in a communication round consists of the pair (S', R') of peer senders and receivers.

2.2 The Hitting-Set Attack

The hitting-set attack (HS-Attack) introduced by [21, 5] is a global passive attack. The goal of the attack is to compute all possible peer recipient sets of a target sender $Alice \in S$ that are called *hypotheses*. Alice's peer recipient set is $\mathcal{H}_A$ and its size is $m = |\mathcal{H}_A|$. We will denote recipients $r \notin \mathcal{H}_A$ by the term *non-peers*. If HS-Attack can find only one hypothesis of size m, then Alice's peer set is uniquely identified. The adversary is interested in Alice's peers, he therefore only observes those pairs (S', R'), where Alice participates as a sender, i.e. $Alice \in S'$. Under this condition we denote the corresponding recipient set R' by the term *observation* $\mathcal{O}$ and the set of observations collected during t rounds is the *observation set* $\mathcal{OS} = \{\mathcal{O}_1, \ldots, \mathcal{O}_t\}$. For each hypothesis $\mathcal{H} \neq \mathcal{H}_A$, it is unlikely that whenever Alice sends a message, also a receiver of $\mathcal{H}$ is contacted. The number of hypotheses therefore decreases with increasing number of observations as illustrated in Example 1.

Example 1. Let $\mathcal{H}_A = \{8, 9\}$ and the observations at time 1,2,3 be $\mathcal{O}_1 = \{8, 5\}$, $\mathcal{O}_2 = \{9, 4\}$, $\mathcal{O}_3 = \{8, 6\}$. Alice contacted peer 8 in the first and third observation and peer 9 in the second observation. At time point 2 the attacker only sees $\mathcal{O}_1, \mathcal{O}_2$, therefore $\mathcal{H} = \{5, 4\}$ is a hypotheses, since these peers could also be contacted by Alice. But at time point 3 $\mathcal{H}$ is excluded, since neither 4 nor 5 is contacted in $\mathcal{O}_3$.

To mount the HS-Attack, the attacker starts with the set $\mathcal{L}_0$ that contains all $\binom{N}{m}$ possible subsets of cardinality m of N recipients, which is called the *hypothesis set*. We assume in this paper that m is know[4], because we are interested in analysing how long Alice can keep a constant set $\mathcal{H}_A$ of m peer partners anonymous. Since Alice has m peer partners, exactly one subset in $\mathcal{L}_0$ is the set of all peer partners of Alice. Let $\{\mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_3, \ldots, \}$ be the observations in the successive communication rounds in which Alice participates. Since Alice has a peer partner in $\mathcal{O}_1$, a set in $\mathcal{L}_0$ that has an empty intersection with $\mathcal{O}_1$ cannot be the set of all peer partners of Alice. Thus upon observing $\mathcal{O}_1$, the attacker obtains a new hypothesis set $\mathcal{L}_1$ by discarding all recipients sets in $\mathcal{L}_0$ that have an empty intersection with $\mathcal{O}_1$. The attacker repeats this process to

[3] A communication round consists of the following events: The Mix node collects messages from a fixed number of distinct senders, and after applying the "Mix" protocol, it forwards the collected messages to their intended recipients.

[4] All attacks shown in this paper are also applicable if m is unknown. See [22] for a justification.

generate hypotheses sets $\mathcal{L}_2, \mathcal{L}_3, \ldots$ after observing recipient sets $\mathcal{O}_2, \mathcal{O}_3, \ldots$ respectively, until the hypothesis set $\mathcal{L}_t$ has only one subset in it. The last remaining subset in the hypothesis set $\mathcal{L}_t$ has to be the set $\mathcal{H}_A$ of all peer partners of Alice, hence the algorithm*fully discloses* Alice's peer set. Note that HS-Attack finds the *unique minimal* hitting set of all observations. A *hitting set* is a set that intersects with all given sets [5]. The hitting set is called *minimal*, if no proper subset of it is a hitting set, otherwise it is called *non-minimal*. All hitting sets addressed in this paper are of size less or equal m. Also note that HS-Attack requires the least number of observations to disclose Alice's peer set under a global passive attacker as proved in [5].

3 Properties of Minimal Hitting Sets

Peers of any hitting sets $\mathcal{H} \neq \mathcal{H}_A$, where $\mathcal{H} \leq m$ are unlikely to be always contacted whenever Alice communicates and this becomes unlikelier, the smaller the size of $\mathcal{H}$ is. After some observations, all hitting sets of size m must therefore be minimal. From now on these minimal hitting sets (of cardinality m) are called *hypotheses* and the *hypothesis set* is the set of all hypotheses. Each non-minimal hitting set $\mathcal{H}$ is a superset of a minimal hitting set $\mathcal{H}'$ of cardinality $m' < m$. Minimal hitting sets therefore represent the common peers of all non-minimal supersets thereof.

A peer can be identified, if it is common to all hitting sets. It is therefore straight forward and without loss of generality to restrict our analysis to minimal hitting sets.

The HS-Attack [21, 5] in Sect. 2.2 does not focus on minimal hitting sets. It simply removes non hitting sets from the set of all $\binom{N}{m}$ possible sets of size m. In contrast to this the ExactHS attack introduced by Pham [23] is the first work that reveals precise structures and quantities of minimal hitting sets. The ExactHS attack is a structured variant of the minimal hitting set attack that requires the same (number of) observations to disclose Alice's peer set as the HS-Attack.

We will extend Pham's work [23] and show how to derive a mathematic model for the evolution of the minimal hitting sets by observations. The obtained model will be elementary, since it enables us to determine the probability, the number and the structure of the minimal hitting sets after any number of observations with respect to the parameters N,b,m of the Mix. In particular we can determine the number of observations, such that a particular number of Alice's peers is disclosed, which is our new metric MTTD.

3.1 Number of Minimal Hitting Sets

The ExactHS attack [23] is a minimal hitting set attack that computes all minimal hitting sets of size lower or equal m with respect to a given observation set (representing the observations of the attacker). It therefore allows us to prove the following claim about the number of minimal hitting sets.

Claim. Let N, b, m be the Mix parameters and $\mathcal{H}_A$ be Alice's peer set of size m. For a given observation set $\mathcal{OS}$, the maximal number of possible minimal hitting sets of cardinality less or equal to m in $\mathcal{OS}$ is b^m. This bound is *sharp*, if $mb \leq N$. For $mb > N$ this is still an upper bound, but it is not sharp.

[5] In our case these sets are the observations $\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_t$.

ExactHS Algorithm. Before the first invocation of Alg. 1 the set of all minimal hitting sets $\mathcal{L}$ and the candidate set $\mathcal{H}$ are empty, and the observation set $\mathcal{OS}$ consists of all observations collected by the attacker. The computation of the minimal hitting sets is initially invoked by calling $\mathit{ExactHS}(\mathcal{OS}, m, \mathcal{H})$. We refer to this observation set by the term *initial observation set*, as $\mathcal{OS}$ will be changed during the processing of ExactHS. In each recursion level $\mathcal{H}$ is extended by exactly one peer r, chosen in Line 7 from a *designated observation* $\mathcal{O} \in \mathcal{OS}$ determined in Line 5, where $\mathcal{OS}$ is the actual observation set. At the invocation of the next recursion level to determine the next peer to be added to $\mathcal{H} \cup \{r\}$ in Line 8, ExactHS is applied to a modified copy of the actual observation set that contains no observations intersecting with $\{r\}$. This step of removing is important to avoid non-minimal hitting sets, as it allows us to focus on adding only those peers to the actual set $\mathcal{H} \cup \{r\}$ that definitively intersect with observations not intersected by $\mathcal{H} \cup \{r\}$. Finally, if Line 2 detects that no observation of the actual observation set remains that is not intersected by $\mathcal{H}$, then $\mathcal{H}$ is a hitting set. In this case it will be added to the set $\mathcal{L}$ in Line 3. After a selection of r has been done in a recursion level, we remove r from all observations of the actual observation set in Line 8 and from the designated observation $\mathcal{O}$ in Line 9. This way the algorithm can repeat the extension of $\mathcal{H}$ with a new peer r not chosen before in Line 7.

Algorithm 1. ExactHS

```
1:  procedure ExactHS($\mathcal{OS}, m', \mathcal{H}$)
2:      if $\mathcal{OS} = \{\}$ then
3:          $\mathcal{L} \leftarrow \mathcal{L} \cup \{\mathcal{H}\}$            ▷ $\mathcal{H}$ is a hitting set, add it to hypothesis set $\mathcal{L}$
4:      else if $m' \geq 1$ then              ▷ add a peer to $\mathcal{H}$, if $\mathcal{H}$ contains less than $m$ peers
5:          choose $\mathcal{O} \in \mathcal{OS}$
6:          while $(|\mathcal{O}| > 0) \wedge (\{\} \notin \mathcal{OS})$ do
7:              choose $r \in \mathcal{O}$                      ▷ $r$ will become element of $\mathcal{H}$
8:              ExactHS $(\mathcal{OS} \setminus \{\mathcal{O}_i \in \mathcal{OS} \mid r \in \mathcal{O}_i\}, m'-1, \mathcal{H} \cup \{r\})$ ▷ select remaining $(m'-1)$ peers of $\mathcal{H}$
9:              $\mathcal{OS} \leftarrow \bigcup_{\mathcal{O}_i \in \mathcal{OS}} \{\mathcal{O}_i \setminus \{r\}\}$     ▷ remove $r$ in all observations of $\mathcal{OS}$
10:             $\mathcal{O} \leftarrow \mathcal{O} \setminus \{r\}$            ▷ do not choose $r$ in this recursion level again
11:         end while
12:     end if
13: end procedure
```

Bound of the Number of Minimal Hitting Sets. ExactHS creates a hitting set $\mathcal{H}$ by starting with an empty set $\mathcal{H} = \{\}$ and adding a recipient to $\mathcal{H}$ in each choice phase represented by the lines 6–11. The number of recursive invocation of the choice phases in Line 8 is restricted by m, since we are interested in computing hitting sets $\mathcal{H}$ with at most m recipients. In each choice phase we only have at most b possible choices of a recipient r_i, because only recipients $r_1, \ldots, r_b$ of a fixed observation $\mathcal{O}$ are selected. From the restriction on the number of recursive invocations of the choice phase and the number of choices in each phase, we can conclude that the algorithm computes at most b^m minimal hitting sets. A formal proof that ExactHS is sound and complete with respect to the computation of all minimal hitting sets can be found in [23].

To show that the bound b^m of the algorithm is tight, we construct m pairwise disjoint observations $\mathcal{O}_1, \ldots, \mathcal{O}_m$, such that $\mathcal{O}_i \cap \mathcal{O}_j = \emptyset$ and $|\mathcal{O}_i| = |\mathcal{O}_j| = b$ for distinct $i, j \in \{1, \ldots, m\}$.

Let us consider a concrete example with the parameters $m = 2, b = 2$, the victim peer set $\mathcal{H}_A = \{1, 2\}$ and the observations $\{1, 3\}, \{2, 4\}$. A short glance shows that there are $b^m = 4$ minimal hitting sets, namely: $\{1, 2\}, \{1, 4\}, \{3, 2\}, \{3, 4\}$.

3.2 Structuring Minimal Hitting Sets

This section shows the classification and quantification of minimal hitting sets introduced by [23].

We partition the minimal hitting sets into $(m + 1)$ disjoint classes $\mathfrak{H}_0, \ldots \mathfrak{H}_m$. A minimal hitting set $\mathcal{H}$ belongs to the class $\mathfrak{H}_i$ (written $\mathcal{H} \in \mathfrak{H}_i$), if and only if it contains exactly $(m - i)$ distinct peer partners of Alice.

For sets A, B and integer i we define $A^{\overline{i}} = \bigcup_{j=0}^{i} A^j$, where A^0 is a neutral element, such that $A^0 \times B^k = B^k$. For example the class $\mathfrak{H}_2$ might contain the (minimal hitting) sets $\mathcal{H}_2 = \{r_{2_1}, r_{2_2}, a_{2_3} \ldots, a_{2_m}\}$ and $\mathcal{H}'_2 = \{r'_{2_1}, a'_{2_3} \ldots, a'_{2_m}\}$, where each r_{i_j} represents a non-peer and each a_{i_k} represents an Alice's peer. All peers within a set must be disjoint.

$$\begin{aligned}
\mathfrak{H}_0 &= \{\mathcal{H}_A\} \\
\mathfrak{H}_1 &\subseteq (R \setminus \mathcal{H}_A)^{\overline{1}} \times \mathcal{H}_A^{m-1} \\
\mathfrak{H}_2 &\subseteq (R \setminus \mathcal{H}_A)^{\overline{2}} \times \mathcal{H}_A^{m-2} \\
&\vdots \\
\mathfrak{H}_m &\subseteq (R \setminus \mathcal{H}_A)^{\overline{m}} \ .
\end{aligned} \tag{1}$$

Bounds of Minimal Hitting Set Classes. The last section derives the bound of b^m for the number of minimal hitting sets. Based on ExactHS (2) represents refined bounds for each of the minimal hitting set classes $\mathfrak{H}_0, \ldots, \mathfrak{H}_m$ as proved in [23].

$$|\mathfrak{H}_i| = \binom{m}{m-i}(b-1)^i = \binom{m}{i}(b-1)^i \ . \tag{2}$$

Note that this bound is again tight and we can use the same construction of m disjoint observations as in Sect. 3.1 to prove its tightness. It is also clear that the sum of the cardinality of each class results in b^m, i.e. $\sum_{i=0}^{m} |\mathfrak{H}_i| = \sum_{i=0}^{m} \binom{m}{i}(b-1)^i = b^m$.

Probability Property of Classes. To model the probability of excluding a particular hypothesis of a class $\mathfrak{H}_i$, we assume that Alice chooses her recipient in each round uniformly distributed from the set of m recipients $\mathcal{H}_A = \{a_1, \ldots, a_m\}$. Similarly the remaining $b - 1$ senders of a batch are assumed to select their receivers uniformly from the set R of N receivers. A (former) hitting set $\mathcal{H}$ is *excluded* by an observation $\mathcal{O}$, if and only if $\mathcal{H}$ does not intersect with $\mathcal{O}$, i.e. if $\mathcal{H} \cap \mathcal{O} = \emptyset$. A hitting set $\mathcal{H}$ is *excludable* with respect to an observation set $\mathcal{OS}$, if and only if $\mathcal{H}$ is a hitting set in $\mathcal{OS}$ and there exist an observation $\mathcal{O} \notin \mathcal{OS}$, such that $\mathcal{H}$ would be excluded by $\mathcal{O}$.

Suppose that a hypothesis $\mathcal{H} \in \mathfrak{H}_i$ is given. According to Pham [24] the probability that this particular hypothesis is excluded by the next observation $\mathcal{O}$ is:

$$p_{inv}(N, b, m, i) = \frac{i}{m}(1 - \frac{m}{N})^{b-1} \ . \tag{3}$$

Thereby the first factor $\frac{i}{m}$ is the probability that Alice chooses to communicate with any of the i recipients not covered by $\mathcal{H}$ in the observation $\mathcal{O}$. The second factor represents the probability that the remaining $(b - 1)$ senders do not choose to contact any of the recipients in $\mathcal{H}$.

3.3 Extensive Hypotheses

Extensive hypotheses combine the knowledge about the minimal hitting sets of size $\leq m$, which are computed by ExactHS with the knowledge about the structure of hypotheses. That way we can predict which hypotheses will be computed in the future.

Table 1 shows the the Minimal hitting sets $\mathcal{M}_i$, the extensive hypothesis set $\mathcal{L}_i$, and the excluded sets that result from analysing the observation set $\mathcal{OS}_i = \{\mathcal{O}_1, \ldots, \mathcal{O}_i\}$.

For $i = 0$ there is no observation and no $\mathcal{M}_0$, but we have knowledge about the initial hypothesis set represented by the classes (1) in $\mathcal{L}_0$. Each element $\mathcal{H} \in \mathcal{L}_i$ is called an *extensive class*. We will address these classes by their order from left to right and from top to bottom, i.e. $\mathcal{H}_u$ is the u-th class in the hypothesis set. An extensive class is *unspecified*, if it contains a variable x (standing for any variable x_u^v with indexes), otherwise it is *specified*. A specified extensive class is called a *specified extensive hypothesis*. The only specified extensive hypothesis in $\mathcal{L}_0$ is $\mathcal{H}_1 = \{1, 2, 3\}$. Each variable x represents any $(b-1)$ *unspecified* non-peers $r \in R \setminus \mathcal{H}_A$, thereby only distinct peers can be assigned to $x_u^v, x_u^w \in \mathcal{H}_u$ for $v \neq w$. Peers that are explicitly mentioned are called *specified*. Newly specified peers are bold highlighted. Note that writing $\mathcal{H} \subseteq \mathcal{L}_0$ would be more appropriate than $\mathcal{H} \in \mathcal{L}_0$, since $\mathcal{H}$ is a class. But for the sake of reducing formalisms and simplifying explanations, we use the element-notation and -operations.

For $i = 3$ we can see that extensive hypotheses also visualise exclusions of implicit hypotheses (e.g. $\{2, 3, 4\}$ and $\{3, 4, 5\}$). A hypothesis is *implicit*, if it has not been computed by ExactHS as a minimal hitting set yet, otherwise it is *explicit*. Finally all extensive hypotheses will be explicit and equal to minimal hitting sets as seen in $i = 4$.

Table 1. Evolution of minimal hitting sets and extensive hypothesis set

i	$\mathcal{O}_i$	Minimal hitting sets $\mathcal{M}_i$	Extensive hypothesis set $\mathcal{L}_i$	Exclusion
0			$\mathfrak{H}_0 : \{1, 2, 3\}$; $\mathfrak{H}_1 : \{1, 2, x_2^1\}, \{1, 3, x_3^1\}, \{2, 3, x_4^1\}$; $\mathfrak{H}_2 : \{1, x_5^1, x_5^2\}, \{2, x_6^1, x_6^2\}, \{3, x_7^1, x_7^2\}$; $\mathfrak{H}_3 : \{x_8^1, x_8^2, x_8^3\}$	
1	$\{1, 4\}$	$\{1\}, \{4\}$	$\mathfrak{H}_0 : \{1, 2, 3\}$; $\mathfrak{H}_1 : \{1, 2, x_2^1\}, \{1, 3, x_3^1\}, \{2, 3, \mathbf{4}\}$; $\mathfrak{H}_2 : \{1, x_5^1, x_5^2\}, \{2, \mathbf{4}, x_6^1\}, \{3, \mathbf{4}, x_7^1\}$; $\mathfrak{H}_3 : \{\mathbf{4}, x_8^1, x_8^2\}$	
2	$\{2, 5\}$	$\{1, 2\}, \{1, 5\}$, $\{4, 2\}, \{4, 5\}$	$\mathfrak{H}_0 : \{1, 2, 3\}$; $\mathfrak{H}_1 : \{1, 2, x_2^1\}, \{1, 3, \mathbf{5}\}, \{2, 3, 4\}$; $\mathfrak{H}_2 : \{1, \mathbf{5}, x_5^1\}, \{2, 4, x_6^1\}, \{3, 4, \mathbf{5}\}$; $\mathfrak{H}_3 : \{4, \mathbf{5}, x_8^1\}$	
3	$\{1, 6\}$	$\{1, 2\}, \{1, 5\}$, $\{4, 2, 6\}, \{4, 5, 6\}$	$\mathfrak{H}_0 : \{1, 2, 3\}$; $\mathfrak{H}_1 : \{1, 2, x_2^1\}, \{1, 3, 5\}$; $\mathfrak{H}_2 : \{1, 5, x_4^1\}, \{2, 4, \mathbf{6}\}$; $\mathfrak{H}_3 : \{4, 5, \mathbf{6}\}$	$\{2, 3, 4\}$, $\{3, 4, 5\}$
4	$\{3, 4\}$	$\{1, 2, 3\}, \{1, 2, 4\}$, $\{1, 5, 3\}, \{1, 5, 4\}$, $\{4, 2, 6\}, \{4, 5, 6\}$	$\mathfrak{H}_0 : \{1, 2, 3\}$; $\mathfrak{H}_1 : \{1, 2, \mathbf{4}\}, \{1, 3, 5\}$; $\mathfrak{H}_2 : \{1, 5, \mathbf{4}\}, \{2, 4, 6\}$; $\mathfrak{H}_3 : \{4, 5, 6\}$	

$\mathcal{L}_i$ is constructed with respect to the observation set $\mathcal{OS}_i = \{\mathcal{O}_1, \ldots, \mathcal{O}_i\}$ and the minimal hitting sets $\mathcal{M}_i$ for $i \geq 1$. Let $\mathcal{H} \in \mathfrak{H}_j$ be an extensive class of size m and $\mathcal{H}^- = \mathcal{H} \setminus \mathcal{H}_A \setminus \{x\}$, then $\mathcal{H} \in \mathcal{L}_i$, if and only if $\mathcal{H}^-$ is a minimal hitting set with respect to $\mathcal{OS}_i' = \{\mathcal{O} \setminus \mathcal{H}_A \mid \mathcal{O} \in \mathcal{OS}_i, \mathcal{O} \cap \mathcal{H} \cap \mathcal{H}_A = \emptyset\}$ [6]. Thus for each $\mathcal{H}^-$, there is a minimal hitting set $\mathcal{H}_i$ with respect to $\mathcal{OS}_i$, such that $|\mathcal{H}_i| \leq m$, $\mathcal{H}^- = \mathcal{H}_i \setminus \mathcal{H}_A$ and $\mathcal{H}_i \subseteq \mathcal{H}$ An extensive class $\mathcal{H}$ that complies to these conditions is called *minimal*, hence an extensive hypothesis set consists of only minimal classes. This defines a surjective

[6] The set $\mathcal{OS}_i'$ results from removing all Alice's peers from those observations in $\mathcal{OS}_i$ that do not contain any of Alice's peers of $\mathcal{H}$.

mapping of the extensive hypothesis set $\mathcal{L}_i$ to the minimal hitting sets with respect to $\mathcal{OS}_i$. We can define $\mathcal{L}_i$ for $i \geq 1$ recursively as follows:

1. Let $\mathcal{L}_i = \{\}$ before the start of its construction below.
2. For each extensive class $\mathcal{H} \in \mathcal{L}_{i-1}$, let $\{r_1, \ldots, r_k\} \subseteq \mathcal{H}$ be the set of all specified peers in $\mathcal{H} \in \mathfrak{H}_j$, where $k \leq m$. Apply either 3. or 4. to each $\mathcal{H}$.
3. If $\{r_1, \ldots, r_k\} \cap \mathcal{O}_i \neq \emptyset$ then add $\mathcal{H}$ to $\mathcal{L}_i$, i.e. $\mathcal{L}_i = \mathcal{L}_i \cup \{\mathcal{H}\}$, because $\mathcal{H}$ is not excluded by $\mathcal{O}_i$.
4. Else if $\{r_1, \ldots, r_k\} \cap \mathcal{O}_i = \emptyset$ and $k < m$ then add for each non-peer $r \in \mathcal{O}_i \setminus \mathcal{H}_A$ the extensive class $\mathcal{H}' = \{r_1, \ldots, r_k, r, x^1, \ldots, x^{m-k-1}\} \in \mathfrak{H}_j$ to $\mathcal{L}_i$, if $\mathcal{H}'$ is a minimal class in $\mathcal{OS}_i$.

We generalise from this example that all extensive classes will become specified apart from the exceptions discussed below. The exclusion probability of a specified extensive hypothesis $\mathcal{H} \in \mathfrak{H}_i$ is exactly $p_{inv}(N, b, m, i)$, even if $\mathcal{H}$ is implicit.

The number of specified extensive hypotheses resulting from $\mathcal{L}_0$ is strictly bounded by b^m. This is due to the fact that $\mathcal{L}_0$ and its extensions are defined according to the classes $\mathfrak{H}_i$ (1) and their class sizes (2).

Exceptions. An *exception* can only arise in point 4. of the computation of $\mathcal{L}_i$ and consists of following cases:

- The extensive class $\mathcal{H}' \in \mathfrak{H}_j$ resulting from specifying a peer in $\mathcal{H} \in \mathfrak{H}_j$ is not minimal in $\mathcal{OS}_i$.
- There are less than $(b-1)$ non-peers in $\mathcal{O}_i$.

We now analyse the effect of exceptions on the number of the sets that will be specified. Let Alice's peers be $\mathcal{H}_A = \{1, 2, 3\}$, $b = 3$ and the considered extensive class be $\mathcal{H} = \{1, 4, x\} \in \mathfrak{H}_2$. If the next observation is no exception (e.g. $\mathcal{O}_i = \{2, 7, 8\}$), then $(b-1) = 2$ specified sets of $\mathfrak{H}_2$ would result from extending $\mathcal{H}$. These sets are $\mathcal{H}' = \{1, 4, 7\}$ and $\mathcal{H}'' = \{1, 4, 8\}$. Assume that $\mathcal{H}''$ is not minimal than only $\mathcal{H}'$ would be the extension of $\mathcal{H}$. Similarly, if the next observation would contain less than $(b-1)$ non-peers, e.g. $\mathcal{O}_i = \{2, 3, 6\}$, respectively $\mathcal{O}_i = \{2, 7\}$. Only one specified set $\mathcal{H}' = \{1, 4, 6\}$ respectively $\mathcal{H}' = \{1, 4, 7\}$ would result from extending $\mathcal{H}$.

Let k be the number of specified peers in $\mathcal{H}$ from point 4., then for each missing non-peer in the next observation $\mathcal{O}_i$, the number of sets that will be specified decreases by at most b^{m-k-1}. The same decrease is caused, if the class $\mathcal{H}'$ resulting from specifying a peer in $\mathcal{H}$ is not minimal in $\mathcal{OS}_i$.

Note that the preconditions for exceptions imply that sets excluded by exceptions are unspecified and implicit before the exclusion. We observe that most extensive hypotheses become specified very fast and logically at least as fast as minimal hitting sets reach the size m. The impact of exclusions by exceptions on the size of the extensive hypothesis set is therefore moderate in comparison to normal (non-exceptional) exclusions. For the sake of simplicity, we only mathematically model the normal exclusion of extensive hypotheses from the initial b^m extensive hypotheses.

The main result of this section is that we can map the inconvenient exclusion process of minimal hitting sets on the exclusion process of the extensive hypothesis set. This again can be simplified to the exclusion process of specified extensive hypothesis set, where the exclusion probability of each set is known. From now on hypotheses and classes are always addressed in terms of specified extensive hypotheses and classes.

4 Modelling Anonymity States

Section 3.3 justified that the evolution of the minimal hitting sets can be modelled by the evolution of the extensive hypothesis set. This section will introduce formulas to describe the deployment of the size and structure of the (extensive) hypothesis set for distinct number of observations and distinct Mix parameters N, b, m. In particular we will answer the following questions:

- How many observation are required to disclose all of Alice's peers?
- How likely is it to find $k \leq m$ of Alice's peers in a random minimal hitting set after t observations?
- What is the average number of observation to disclose at least one peer of Alice?

4.1 Full Disclosure

The *full disclosure* of Alice's peer set is the unambiguous identification of all of Alice's peer recipients, i.e. the identification of $\mathcal{H}_A$.

Mean Number of Minimal Hitting Sets. In this section, we derive closed formulas for the mean number of hypotheses after t observations for distinct classes.

Let V_i be a random variable, where $V_i = 1$ is the event that a particular hypothesis $\mathcal{H}$ of the class $\mathfrak{H}_i$ remains valid after t observations, while $V_i = 0$ denotes the inverse event. The probability of $V_i = 1$ corresponds to t stochastically independent Bernoulli trials, where the outcome of each of the t trials shows that the minimal hitting set remains valid. Thereby a Bernoulli trial corresponds to the outcome, whether $\mathcal{H}$ remains valid at the next collected observation. Since each observation appears stochastically independently from those in the past and in the future, we have a natural mapping of the "remaining valid event" of $\mathcal{H}$ on the independent Bernoulli trials, hence:

$$P(V_i = 1) = [P(\mathcal{H} \textit{ remains valid at next observation})]^t$$

is the probability that $\mathcal{H}$ remains a hypothesis after t observations, which is by (3) exactly $(1 - p_{inv}(N, b, m, i))^t$.

Let $\mathfrak{H}_i = \{\mathcal{H}_1, \ldots, \mathcal{H}_{|\mathfrak{H}_i|}\}$ be a minimal hitting set class containing only hypotheses and V_{i_j} be the event that the hypothesis $\mathcal{H}_j \in \mathfrak{H}_i$ remains valid after t observations. The expectation E of the number of hypotheses in $\mathfrak{H}_i$ after t observations is thus the expectation of the convolution of the random variables $V_{i_1}, \ldots, V_{i_{|\mathfrak{H}_i|}}$. This expectation is represented by (4). Thereby we benefit from the additivity of the expectation function by splitting the complex expectation on the left side to a sum of expectation of single V_{i_j} events on the right side. The probability of the outcome $P(V_{i_j} = 1)$ is identical for each fixed hypothesis $\mathcal{H}_j \in \mathfrak{H}_i$ (i.e. $P(V_{i_j} = 1) = P(V_i = 1)$), hence the right side of the former equation can be simplified to (5).

$$E(V_{i_1}, \ldots, V_{i_{|\mathfrak{H}_i|}}) = \sum_{j=1}^{|\mathfrak{H}_i|} E(V_{i_j}) \tag{4}$$

$$= \sum_{j=1}^{|\mathfrak{H}_i|} P(V_{i_j} = 1)$$

$$= |\mathfrak{H}_i| P(V_i = 1) \ . \tag{5}$$

Note that the events V_{i_j}, V_{i_k} for $i, k \in \{1, \dots |\mathfrak{H}_i|\}$ and $j \neq k$ are not stochastically independent, hence the probability $P(V_{i_j}) = P(V_i)$ respectively $P(V_{i_k}) = P(V_i)$ only holds, if we consider single events, as on the right side of the equation below.

To clarify that (5) depends on the parameter N, b, m and t we use the more elaborate formulation:

$$E_{|\mathfrak{H}_i|}(N, b, m, t) = |\mathfrak{H}_i|(1 - p_{inv}(N, b, m, i))^t = \binom{m}{i}(b-1)^i(1 - \frac{i}{m}(1 - \frac{m}{N})^{b-1})^t \quad (6)$$

for the expected number of remaining hypotheses of class $\mathfrak{H}_i$.

Formula (4) can be easily extended to cover the mean number of observations for any combination of classes including the consideration of all classes. The expectation of the remaining hypotheses with respect to the initial hypothesis set $\mathfrak{H}$ is:

$$\begin{aligned} E_{|\mathfrak{H}|}(N, b, m, t) &= \sum_{i=0}^{m} \binom{m}{i}(b-1)^i(1 - \frac{i}{m}(1 - \frac{m}{N})^{b-1})^t \\ &\leq ((b-1)e^{-\frac{t}{m}(1-\frac{m}{N})^{b-1}} + 1)^m \ . \end{aligned} \quad (7)$$

Time to Reduce Hypothesis Set to a Threshold. The expectations $E_{|\mathfrak{H}_i|}$ and $E_{|\mathfrak{H}|}$ of the number of hypotheses after t observations can be easily reformulated to derive the number of observations, such that a hypotheses remains on average.

By a transformation of (6), where a denotes the left side of the equation, we obtain:

$$t_{\mathfrak{H}_i} = \frac{\ln a - \ln \binom{m}{i} - i \ln (b-1)}{\ln (1 - \frac{i}{m}(1 - \frac{m}{N})^{b-1})} \quad \text{for } a > 0 \ . \quad (8)$$

This equation represents the number of observations, such that at most a hitting sets remain on average in the class $\mathfrak{H}_i$ for $i \geq 1$.

Similarly we reformulate (7) to obtain the number of observations $t_{\mathfrak{H}}$, such that there are on average less than a minimal hitting sets left from the initial hypothesis set $\mathfrak{H}$. Alice's peer set $\mathcal{H}_A$ always remains in $\mathfrak{H}$, therefore $a > 1$.

$$t_{\mathfrak{H}} \leq \frac{m(\ln (b-1) - \ln (a^{1/m} - 1))}{(1 - \frac{m}{N})^{b-1}} \quad \text{for } a > 1 \ . \quad (9)$$

Comparison to Simulation. This section visualise the precision of the function $t_{\mathfrak{H}}$ of Sect. 4.1 by comparing it with the mean time of full disclosure obtained by our hitting set simulations. The simulation applies the HS-Attack on simulated observations, until Alice's peer set can be uniquely identified. This simulation is run several times to obtain a confidence interval of 95% on the mean number of observations to identify Alice's peer set. The observations are generated under the assumption of a uniformly distributed communication of Alice and the other senders. That is Alice chooses one of her m peers with probability $\frac{1}{m}$ and each of the other senders chooses its peer from all N receivers with the probability $\frac{1}{N}$ at each round. This distribution complies to the distribution used in our formulas.

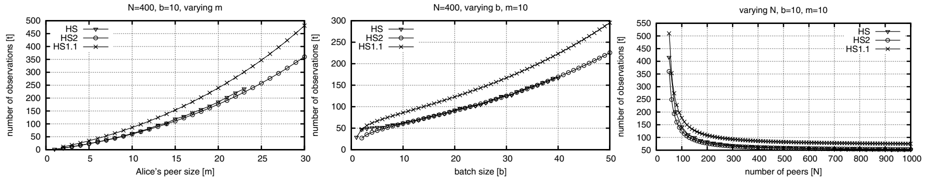

Fig. 2. Number of observations: Full disclosure by simulation (HS), reduction of hypothesis set to size below 2 (HS2) and below 1.1 (HS1.1)

The plots in Fig. 2 compare the mean number of observations for full disclosure obtained by the simulation (HS) with the number of observations to reduce the initial hypothesis set to a cardinality less than 2 (HS2) respectively less than 1.1 (HS1.1) using (9). The y-axes of the plots shows the number of observations, while the x-axes vary one of the parameters N,b, or m.

Note that the mean number of observations for full disclosure is not necessary equal to the number of observation to reduce the hypothesis set to a particular size, although these two values are strongly related to each other. Let μ_{dis} be the mean number of observations for the full disclosure, then the mean number of hypotheses after μ_{dis} observations is obviously larger than 1. Depending on the variance of the number of hypotheses around μ_{dis}, more than μ_{dis} observations are necessary to keep the number of hypotheses closed to 1. This is shown by the Fig. 2. The HS2 curve is almost identical to the HS curve, whereas the HS1.1 curve is noticeably above the HS curve.

Parallel to us [25] suggested a “lower bound” t^* for the mean number of observations for full disclosure using the same Mix and attacker model. They compute for each of the $\binom{N}{m}$ “normal” (not extensive) hypotheses the probability that it is excluded after t observations and claim that t^* is the lower bound of the time when only one hypothesis remains. Figure 3 shows that t^* is far away from being a lower bound. Firstly, there are at most $\binom{N}{m} - \binom{N-b}{m}$ “normal” hypotheses after the first observation due to the Mix model, but this restriction is not in their mathemtic model and causes a significant overestimation. Secondly, even if this would be corrected, the evolution of the “normal” hypothesis set depends on the distribution of the hypotheses' structures remaining after the first observation and those succeeding that. This is not mathematically modeled and might be very difficult to do.

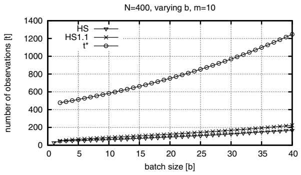

Fig. 3. Comparison: Simulation (HS), reducing hypothesis set below 1.1 (HS1.1), t^* (t*)

4.2 Partial Disclosure

The *partial disclosure* is the unambiguous identification of a subset $\mathcal{H}_{A'} \subseteq \mathcal{H}_A$ of Alice's peer set. The full disclosure in Sect. 4.1 is a special case of the partial disclosure.

Probability to Identify k Particular Peers. The probability to identify k particular peers $\mathcal{H}_{A'} \subseteq \mathcal{H}_A$ of Alice after at most t observations is the probability that all hypotheses are excluded that do not contain all of these $k = |\mathcal{H}_{A'}|$ peers after at most

t observations. This probability is a discrete distribution with respect to t and we will address it by the term f_{id}. The probability to exclude a particular hypothesis depends on its class, therefore we will first determine the number of hypotheses of each class $\mathfrak{H}_i$ that have to be excluded.

Number of Exclusions in a Class. We remember from (2) that the size of $\mathfrak{H}_i$ is $|\mathfrak{H}_i| = \binom{m}{i}(b-1)^i$. In this class i of the m peers of Alice are replaced by non-peers. Therefore $\binom{m-k}{i}(b-1)^i$ is the number of hypotheses in the class $\mathfrak{H}_i$, where the k of Alice's peers $\mathcal{H}_{A'}$ are not replaced by non-peers. The number of hypotheses in $\mathfrak{H}_i$ that have to be excluded to enable the identification of $\mathcal{H}_{A'}$ is therefore:

$$exNo_i(b,m,k,i) = \left(\binom{m}{i} - \binom{m-k}{i}\right)(b-1)^i \ . \tag{10}$$

Note that we distinguish between the to be excluded hypotheses with respect to their class membership, because the probability to exclude a hypothesis depends on its membership as shown in Sect. 3.2. Also note that we only know the probability with respect to the exclusion of a single hypothesis. If two hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$ are considered, then there could be a stochastic dependency between them, i.e. if $\mathcal{H}_1$ is excluded, then $\mathcal{H}_2$ might be excluded, too. For that reason we make the simplifying assumption that all minimal hitting sets are stochastically independent. This enables us to unrestrictedly apply p_{inv} to describe the exclusion of hypotheses. The following equation derives the distribution f_{id} with respect to the parameters N, b, m, t and the number $k = |\mathcal{H}_{A'}|$ of Alice's peers that should be identified.

$$\begin{aligned} f_{id}(N,b,m,k,t) = &\prod_{i=1}^{m-k} (1-(1-p_{inv}(N,b,m,i))^t)^{(\binom{m}{i}-\binom{m-k}{i})(b-1)^i} \\ &\prod_{i=m-k+1}^{m} (1-(1-p_{inv}(N,b,m,i))^t)^{\binom{m}{i}(b-1)^i} \ . \end{aligned} \tag{11}$$

Probability to Identify at Least *k* Peers. Based on the function f_{id} of the last section, we will derive the probability distribution $f_{id_{any}}$ that at least k of Alice's peers can be identified after at most t observations. In contrast to the previous section we are not focusing on disclosing particular peers, but on the probability to disclose a certain number of peers.

Let Y^k be a random variable denoting the event that particular k peers of Alice's peer set $\mathcal{H}_A$ are identified, i.e. $Y^k = 1$ if the designated peers are identified else $Y^k = 0$ for the inverse case. To simplify the notation we will abbreviate the probability $P(Y^k = 1)$ by the term $P(Y^k)$.

Let $Y_1^k, \ldots, Y_{\binom{m}{k}}^k$ be $\binom{m}{k}$ distinct random variables. Each of this variable represents the event that distinct subsets of $\mathcal{H}_A$ of cardinality k are identified. In order to compute the probability that at least k of Alice's peers can be disclosed, we have to determine the probability that any of these Y_i^k events, for $i \in \{1, \ldots, \binom{m}{k}\}$ takes place. Thereby it would be imprecise to simply sum up the probabilities $P(Y_i^k)$ for $i \in \{1, \ldots, \binom{m}{k}\}$, because the events Y_i^k are not stochastically independent. We can solve this problem by applying the inclusion-exclusion-formula:

$$\begin{aligned} P(Y_1^k \vee \ldots \vee Y^k_{\binom{m}{k}}) &= P(Y_1^k) + \ldots + P(Y^k_{\binom{m}{k}}) \\ &\quad - P(Y_1^k, Y_2^k) - \ldots - P(Y^k_{\binom{m}{k}-1}, Y^k_{\binom{m}{k}}) \ldots + \ldots - \ldots \ . \end{aligned} \tag{12}$$

Assume that $\{a_{i_1}, a_{i_2}\}$ and $\{a_{j_1}, a_{j_2}\}$ are those peers that are identified by the event Y_i^k respectively Y_j^k (for $k = 2$). The above term $P(Y_i^k, Y_j^k)$ is the probability that all peers of the joint set $\{a_{i_1}, a_{i_2}, a_{j_1}, a_{j_2}\}$ are identified. Let us denote the joint event by the term $Y^{k'}$, where $k' = |\{a_{i_1}, a_{i_2}, a_{j_1}, a_{j_2}\}| \leq 2k$, then $P(Y_i^k, Y_j^k) = P(Y^{k'})$ can be computed by (11). It is also obvious that this transformation can even be applied to an arbitrary number of joints of events, i.e. we can transform $P(Y_1^k, \ldots, Y_z^k)$ to $P(Y^{k'})$ for any $z \geq 1$ accordingly.

The next formula is an elaborate formulation of (12) for the special case of $k = 1$. It is the probability to identify at least one of Alice's peers after at most t observations.

$$f_{id_{any}}(N, b, m, t, 1) = \sum_{s=1}^{m} \left((-1)^{s-1} \binom{m}{s} f_{id}(N, b, m, s, t) \right) \ . \tag{13}$$

The general probability distribution for arbitrary values of k, where $k \leq m$ is the least number of peers that are to be disclosed is:

$$f_{id_{any}}(N, b, m, t, k) = \sum_{i=1}^{\binom{m}{k}} (-1)^{i-1} \sum_{j_1=1}^{\binom{m}{i}-(i-1)} \cdots \sum_{j_i=j_{i-1}+1}^{\binom{m}{i}-(i-i)} f_{id}(N, b, m, |\bigcup_{z=1}^{k} Y_{j_z}|, t) \ ,$$

where $\bigcup_{z=1}^{k} Y_{j_z}$ is the union of the set of peers identified by each Y_{j_z}.

Given the distribution $f_{id_{any}}(N, b, m, t, k)$, the probability that at least k peers can be identified after exactly t observation is:

$$p_{id_{any}}(N, b, m, t, k) = f_{id_{any}}(N, b, m, t, k) - f_{id_{any}}(N, b, m, t-1, k) \ .$$

Mean Time to Deanonymization. We are now able to provide the first existing formula to compute the mean number of observations to unambiguously identify at least k of Alice's peer, which we call MTTD-k.

$$E_{id_{any}}(N, b, m, t, k) = \sum_{t=1}^{\infty} t \, p_{id_{any}}(N, b, m, t, k) \ . \tag{14}$$

Note that in particular $E_{id_{any}}(N, b, m, t, 1)$, which is the mean number of observations needed to identify at least one of Alice's peers (MTTD-1) should be considered as a more appropriate measurement of the lower bound of the anonymity provided by Mix systems. This is justified by the fact that MTTD-1 measures the time point, when the attacker gains the first unambiguous information about Alice's communication partners and thus breaks the anonymity function of the Mix. In contrast to this, full disclosure,

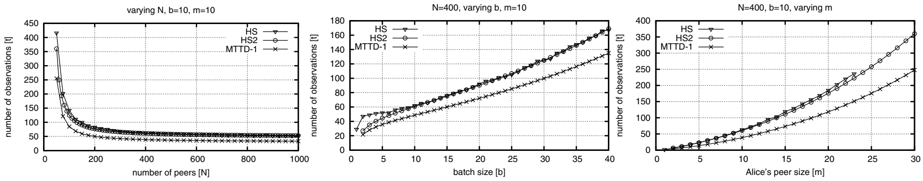

Fig. 4. Disclosure of at least one peer (MTTD-1), simulated full disclosure (HS), reduction of hypothesis set size below 2 (HS2)

or the number of observations to reduce the hypothesis set below a particular size a can not expose this threat.

Figure 4 compares the expected number of observations to disclose at least one peer (MTTD-1) by using $E_{id_{any}}(N, b, m, t, 1)$ with the simulation result for the mean number of observations for full disclosure (HS) and the mean number of observation to reduce the hypothesis set to a size below 2 (HS2) computed by (9). The comparison is with respect to different parameters N, b, m. We can see that the partial disclosure (MTTD-1) appears noticeable earlier than full disclosure (HS) and before the hypothesis set is reduced to a size below 2. This difference increases, the more observations are required for full disclosure and shows that full disclosure alone is insufficient to measure anonymity.

4.3 Beyond Unambiguous Identification of Peers

Our model also enables us to consider the number of Alice's peers contained in any computed minimal hitting set after a particular number of observations.

Figure 5 plots the number of observations to reduce each minimal hitting class $\mathfrak{H}_i$ to a size less than a by using (8). The figure shows this for $a = 1$ by the HS1 curve and for $a = 0.1$ by the HS0.1 curve for the Mix parameters $N = 400, b = 10, m = 10$.

We can see that minimal hitting set classes $\mathfrak{H}_i$ containing less of Alice's peers are reduced earlier than those containing more of Alice's peers. Thus after a particular number of observations t, the number of hypotheses containing less than k Alice's peers are negligible, since $E_{|\mathfrak{H}_i|}(t) < a$ for $i > (m - k)$. In particular our plot shows that after about $t = 40$ observations, the attacker will unlikely find a minimal hitting set containing less than 7 of Alice's peers. Thus any minimal hitting set computed by the attack contains with a high probability at least 70% of the peers of Alice. If we assume that minimal hitting sets are excluded stochastic independently from each other, then the probability to find at least k of m peers of Alice after at most t observations is:

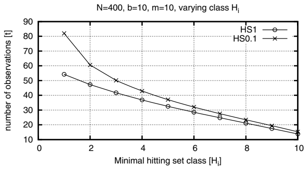

Fig. 5. No. of observ. to reduce size of $\mathfrak{H}_i$ below 1 (HS1) and 0.1 (HS0.1).

$$f_{id_k}(N, b, m, k, t) \geq \prod_{i=m-k-1}^{m} (1 - (1 - p_{inv}(N, b, m, i))^t)^{|\mathfrak{H}_i|} \ .$$

5 Conclusions

In this work, we investigated the fundamental structures for anonymity that we identified as the hypothesis set. The analysis of the hypothesis set is made under the assumption of a uniformly distributed communication of the senders and of static peer sets. This assumption is chosen in a way, such that we obtain a conservative consideration of anonymity, which can be considered as a lower bound of the anonymity provided by Mixes in the real world.

Based on the ExactHS [23], we derived a comprehensive mathematic model to probabilistically describe the inherent properties of the hypothesis set in detail. In particular we estimated in Sect. 4 the size of the hypothesis set, the structure of hypotheses in it, the size of those structures and the probability that particular hypotheses are included in it, with respect to the parameters N,b,m at any number of observations. This information enabled a fine granular measurement of anonymity that also measures those protection states before the point of full disclosure. In particular MTTD-k introduced in Sect. 4.2 determined the mean number of observations to disclose from one to all Alice's peers. The evaluations of MTTD -1 (see Fig. 4) showed that the first unambiguous knowledge about one of Alice's peers can be gained noticeably before full disclosure. It is therefore not sufficient to solely consider full disclosure (which is the focus of all existing hypothesis set based approaches e.g. [5, 23, 25]) for anonymity measurement.

Furthermore, our model even enabled an analysis granularity beyond the scope of unambiguous identification of peers. This is shown in Sect. 4.3, which provided the probability to find a certain number of Alice's peers in randomly computed hypotheses. This insight opens the door for a new refined metric, which also covers unambiguous information in the anonymity consideration.

We also showed that our mathematic model and the resulting measurements were precise and meaningful by comparison to simulations. All results were in reasonable scopes and reflected the right relations to the Mix parameters N, b, m and the number of observations t.

The elementary model and analyses of our work are important for designers and users of Mix networks. It enables Alice to estimate how much information she is going to leak about her peers with each of her communications. That way she knows when to stop communicating, or to add dummy traffic to remain information theoretic anonymous, such that even attackers with unlimited computing power cannot reveal her peers.

In the future we intend to integrate the leakage of unambiguous and ambiguous information about Alice's peers in one metric. In conjunction with this, we will analyse the uncertainty caused by dummy traffic. Finally we will refine our model to expand the analysis beyond the uniform communication assumption to get closer to the real world.

References

[1] Chaum, D.L.: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. Communications of the ACM 24(2), 84–88 (1981)

[2] Pfitzmann, A., Köhntopp, M.: Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology. In: Federrath, H. (ed.) Designing Privacy Enhancing Technologies. LNCS, vol. 2009, pp. 1–9. Springer, Heidelberg (2001)

[3] Eusgeld, I., Freiling, F.C., Reussner, R. (eds.): Dependability Metrics. LNCS, vol. 4909. Springer, Heidelberg (2008)

[4] Shannon, C.E.: Communication theory of secrecy systems. Bell Syst. Tech. J. 28, 656–715 (1949)

[5] Kesdogan, D., Agrawal, D., Pham, V., Rauterbach, D.: Fundamental Limits on the Anonymity Provided by the Mix Technique. In: IEEE Symposium on Security and Privacy (May 2006)

[6] Clauß, S., Schiffner, S.: Structuring Anonymity Metrics. In: DIM 2006: Proceedings of the second ACM workshop on Digital identity management, pp. 55–62 (2006)

[7] Deng, Y., Pang, J., Wu, P.: Measuring Anonymity with Relative Entropy. In: Dimitrakos, T., Martinelli, F., Ryan, P.Y.A., Schneider, S. (eds.) FAST 2006. LNCS, vol. 4691, pp. 65–79. Springer, Heidelberg (2007)

[8] Díaz, C., Seys, S., Claessens, J., Preneel, B.: Towards Measuring Anonymity. In: Dingledine, R., Syverson, P.F. (eds.) PET 2002. LNCS, vol. 2482, pp. 54–68. Springer, Heidelberg (2003)

[9] Edman, M., Sivrikaya, F., Yener, B.: A Combinatorial Approach to Measuring Anonymity, 356–363 (2007)

[10] Serjantov, A., Danezis, G.: Towards an Information Theoretic Metric for Anonymity. In: Dingledine, R., Syverson, P.F. (eds.) PET 2002. LNCS, vol. 2482, pp. 259–263. Springer, Heidelberg (2003)

[11] Tóth, G., Hornák, Z., Vajda, F.: Measuring Anonymity Revisited. In: Proceedings of the Ninth Nordic Workshop on Secure IT Systems, pp. 85–90 (November 2004)

[12] Zhu, Y., Bettati, R.: Anonymity vs. Information Leakage in Anonymity Systems. In: ICDCS 2005: Proceedings of the 25th IEEE International Conference on Distributed Computing Systems, pp. 514–524 (2005)

[13] Clarkson, M.R., Myers, A.C., Schneider, F.B.: Belief in Information Flow. In: Proceedings of the 18th IEEE workshop on Computer Security Foundations, pp. 31–45 (2005)

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

[15] Mantel, H.: A Uniform Framework for the Formal Specification and Verification of Information Flow Security. PhD thesis, Universität des Saarlandes (July 2003)

[16] Schneider, S., Sidiropoulos, A.: CSP and Anonymity. In: Martella, G., Kurth, H., Montolivo, E., Bertino, E. (eds.) ESORICS 1996. LNCS, vol. 1146, pp. 198–218. Springer, Heidelberg (1996)

[17] Halpern, J.Y., O'Neill, K.R.: Anonymity and Information Hiding in Multiagent Systems 13, 483–514 (2005)

[18] Hughes, D., Shmatikov, V.: Information Hiding, Anonymity and Privacy: a Modular Approach. J. Comput. Secur. 12, 3–36 (2004)

[19] Padlipsky, M.A., Snow, D.W., Karger, P.A.: Limitations of End-to-End Encryption in Secure Computer Networks. Technical Report ESD-TR-78-158 (August 1978)

[20] Pfitzmann, A.: Diensteintegrierende Kommunikationsnetze mit teilnehmerüberprüfbarem Datenschutz. Informatik-Fachberichte, vol. 234 (1990)

[21] Kesdogan, D., Pimenidis, L.: The Hitting Set Attack on Anonymity Protocols. In: Fridrich, J. (ed.) IH 2004. LNCS, vol. 3200, pp. 326–339. Springer, Heidelberg (2004)

[22] Kesdogan, D., Agrawal, D., Penz, S.: Limits of Anonymity in Open Environments. In: Petitcolas, F.A.P. (ed.) IH 2002. LNCS, vol. 2578, pp. 53–69. Springer, Heidelberg (2003)

[23] Pham, V.: Analysis of the Anonymity Set of Chaumian Mixes. In: 13th Nordic Workshop on Secure IT-Systems (October 2008)

[24] Pham, D.V.: Analysis of Attacks on Chaumian Mixes (Analyse von Angriffen auf Chaummixen). Master's thesis, RWTH-Aachen (April 2006)

[25] O'Connor, L.: Entropy Bounds for Traffic Confirmation. Cryptology ePrint Archive (2008), `http://eprint.iacr.org/2008/`