Discussing Anonymity Metrics for Mix Based Anonymity Approaches

Dang Vinh Pham[1] and Joss Wright[1]

Siegen University, Siegen, Germany
pham@fb5.uni-siegen.de, wright@fb5.uni-siegen.de

Abstract. Today the Chaumian Mix idea is not only an academic approach for anonymous communication, it is also a practical system with many variants used in real-world scenarios. It is therefore important to find an appropriate measure for the anonymity provided by these approaches. Many measurement approaches have been proposed that consider only the static state of the system without accounting for past and future information. Still other measurements evaluate only statistics. These measurements have in common that they do not measure when the anonymity function of the system is broken. Inspired by the idea of unicity distance in cryptography, and the mean time to failure in dependable systems, we believe that measuring the point at which the system fails to hide the relation between a sender and a receiver is a more appropriate measure of its anonymity. In this paper, we discuss our arguments with respect to existing measurement approaches.

1 Introduction

Confidentiality of communication relations is a core requirement for many interactions in the Internet, for example in healthcare systems, electronic voting and commerce. The most widely used and practically applicable system for confidential communication is the Mix, introduced by Chaum in 1981[1]. These systems are collectively referred to as *anonymity systems*. In practice they can be standalone, or can appear as the network anonymisation layer of other privacy-preserving systems such as Idemix[2].

The existing Mix approaches, including the pool mix, threshold mix and stop-and-go Mix, have their origins in the basic concept proposed by Chaum. The underlying idea of these systems is to embed a single user within a set of users such that the actions of that user is not identifiable within the set. This set is called the *anonymity set*, and the embedding function is provided by the Mix.

Inspired by this approach, many variants of Mix systems have been proposed that overcome limitations in the original design. Other systems, such as well know onion-routing approach[3] can also be considered to arise from the basic idea of Chaum, but relax the embedding by removing some security functions provided by the Chaumian Mix.

A natural interest of users and designers of a system is to know the strength of the system in anonymising users. We suggest that a measurement of the strength should refer to the following questions:

– How long does it take on average to reveal a communication relation?
– How hard it is to break the anonymity function?

The first question can be considered as taking an information theoretic view, while the second is complexity theoretic. In the first case we are inspired by the mean time to failure in dependable systems, and by Shannon's unicity distance[4] in cryptography. In the second model the unicity distance measures the average number of bits that an attacker must learn to uniquely identify a message by by examination of the cipher text. This number represents how many cipher text characters must be collected by the attacker in order to identify the message unambiguously. The unicity distance is an information theoretic measurement of the strength of a cryptography system with respect to:

– the structure of the system, including the domain of the plaintext, ciphertext and the key.
– the redundancy of the source language.

Applying this concept to anonymity we wish to find the mean number of observations that must be made by an attacker in order to uniquely identify a communication relationship between two parties. This measurement is with respect to:

– the structure of the system, including the domain of senders and receivers, and the size of the anonymity set generated by the system.
– the redundancy of communication by the user.

The second complexity theoretic question results from the idea of security measurement in public key encryption systems. A system is considered to be practically secure if it is *computationally hard* to break. In the case of an anonymity system we are interested in how much computational effort is required to uniquely link a sender to its recipient, given that it is possible in an information theoretic sense.

Our idea is supported by the existing research based on the hitting-set attack[5, 6, 7], an algorithm that can be applied against Mixes under the assumption of a passive attacker that only observes the input and output of the system. The hitting-set attack enables us to see how the set of possible communication partners of a user decreases information is gained from observing the sets of possible communications relationships. We suggest to model this process of information gain to obtain an appropriate information theoretic measurement for the strength of the system. A complexity theoretic measurement then naturally results from analysing the complexity of the hitting-set attack.

Since all other Mix based approaches are variants of the Chaumian Mix, we believe that our measurement idea can be further adapted to the other variants. In contrast, we argue that approaches that do not incorporate the knowledge of the attacker like [8, 9], or that rely only on statistical evaluations, such as the statistical disclosure attack[10] are less appropriate as an anonymity measurement.

We will now define the basic concepts on which later work in this paper relies.

2 The Mix Model

We introduce here a formal definition of the Mix model as described in [6], based on the original design introduced in [1].

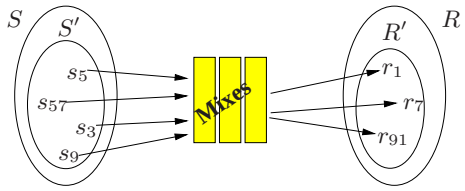

Fig. 1. Formal model

Figure 1 shows the basic components of this technique, consisting of a set of senders, S; a set of recipients, R; and a Mix node. All senders in S communicate with the Mix, and the Mix itself communicates with all recipients in R via a network of secure and reliable channels. A secure reliable channel ensures against loss or duplication of transmitted messages and guarantees authenticity, integrity and confidentiality of transmitted messages. The users and the Mix transmit messages using the following protocols:

User Protocol Users prepare messages by padding them to constant length, with longer messages split over multiple chunks. Each message is then encrypted twice with one time pads. The first layer of encryption uses a shared secret between the sender and the intended recipient, the second layer employs a shared secret between the sender and the Mix. The doubly-encrypted messages are then sent to the Mix.

Mix Protocol A Mix collects a *batch* of b messages from distinct users and then decrypts these messages. The decrypted messages are then output from the mix according to a non-order preserving function such as random delay or lexicographical sorting. The output is broadcast to all recipients. Incoming packets are also compared against formerly received messages in order to reject any duplicates.

The Mix technique described above can perfectly hide the communication relationships between senders and recipients of messages from everybody but the Mix and message senders. If the protocol is applied in fixed time slots, with each user required to supply a fixed number of messages per batch, the act of sending or receiving can itself be hidden[11, 1, 12]. Pfitzmann[12] states that this Mix technique provides information-theoretic anonymity and unobservability based on complexity theoretic secure cryptography.

2.1 The Pure Mix Technique

The "perfect" anonymity solution discussed above makes use of dummy messages and a broadcast function. Even though this solution can provide perfect anonymity it is not practical in large networks such as the Internet, as justified in [6]. As a consequence, most current implementations and solutions use a variant of the perfect Mix solution

by omitting either dummy messages or the broadcast function. We refer to these more practical approaches as *pure* Mix techniques.

Since the pure Mix is not information theoretically secure, the question arises: how much information is leaked by such a technique? Equivalently: how much information must be obtained to break the anonymity provided by these techniques? In order to measure the non-protocol specific anonymity provided by this technique we assume a *passive attacker*. This attacker gains knowledge only by observing sets of senders and receivers at each round of communication. To give the attacker the ability to observe the entire set of senders and receivers we additionally assume that he is *global*. We thus assume a common attacker model for analysis of anonymity systems: the *global passive attacker*.

We consider all approaches in this paper in the context of the pure Mix, which is the basic model underlying the threshold Mix. Our model is general as other Mix techniques, such as the pool Mix, can easily be modelled by a threshold Mix as shown in [13].

We will use the following formal model of a pure Mix and information leakage for our analysis:

Formal Model of the Pure Mix Technique

- A communication system consists of a set of *senders*, S; a set of *recipients*, R; and a Mix node as shown in Figure 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*, consisting of the collection of messages by the Mix and the forwarding of messages to their recipients via the Mix protocol, 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. In this model we do not consider dummy messages to hide sending and receiving actions.
- 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.

3 Unicity Distance

3.1 Unicity Distance and Secrecy

In 1949, in order to measure the theoretical secrecy of cryptographic algorithms, Shannon introduced the concept of *unicity distance*[4, 14]. In the work, a cipher is considered abstractly as a set of mappings from a plain text domain, $\mathcal{D}_{plain}$, to a cipher text domain, $\mathcal{D}_{cipher}$. Each individual mapping is determined by a cryptographic key, and the

set of possible mappings between $\mathcal{D}_{plain}$ and $\mathcal{D}_{cipher}$ is determined by the key space $\mathcal{D}_{key}$.

In this measurement, Shannon considers the possible plaintext messages that could be mapped to a given ciphertext. It is assumed that the attacker is passive and can observe the cipher text, and that he has knowledge about the domains $\mathcal{D}_{plain}$, $\mathcal{D}_{cipher}$ and $\mathcal{D}_{key}$. A system is defined as *perfectly secret* if each cipher text can equally result from any possible plain text in $\mathcal{D}_{plain}$. In such a case, the ciphertext provides the attacker with no information concerning the plain text and he is thus unable to identify unambiguously the plaintext. The canonical example of such a perfectly secret system is the One-Time-Pad.

Since the overhead for perfect secrecy is very high, practical systems rarely provide this property. Practical ciphers therefore typically leak some information about the plaintext in the ciphertext. As the size of the ciphertext increases, the uncertainty about the original plain text correspondingly decreases. This results from the redundant information in human languages, and the restricted key space that is a concession to performance over security in practical systems.

A language such as English has a particular syntactic structure, and particular frequencies of distinct letters in a sentence such that words and letters are repeated in a specific order. This characteristic of the plaintext is reflected by the ciphertext if the set of possible mappings from plaintext to ciphertext is too small to mask the characteristics by randomness. In particular, if the set of mapping functions is small and the cipher text is large, it is unlikely that there are mappings that map distinct meaningful plaintext messages to the same cipher text. In such a case the original message of the cipher text can theoretically be revealed unambiguously.

The unicity distance, ud, measures the amount of information in bits that an attacker must collect in order to identify unambiguously the original plaintext. This corresponds to the length of ciphertext that an attacker has to observe. Therefore, if an attacker observes less than ud bits of ciphertext then the cryptographic algorithm provides information theoretic secrecy: it is not possible unambiguously to identify the original plaintext message. If the attacker observes more than ud bits, the system is not information theoretically secret and it therefore theoretically possible to reveal the original message from the ciphertext. This can be achieved, for example, by a brute force attack that tries all inverse mappings from the ciphertext to the plaintext.

3.2 Secrecy by Computational Complexity

A system that is not information theoretically secret is not necessary insecure. Public key systems do not typically provide information theoretic secrecy, but they can be considered secure with respect to the amount of computational power to break them. A second way to measure the strength of a cryptographic system is therefore to measure the complexity of breaking the system given that it is not information theoretically secret. In order to achieve this, an attacker must know an algorithms that is capable of breaking the cipher under consideration.

3.3 Unicity Distance and Anonymity

We identify parallels between the anonymity measurement of Mix based systems and the secrecy measurement in cryptography. Without restriction of generality we can assume that messages and ciphertexts consist of letters taken from the latin alphabet. The domain of the plaintext alphabet corresponds to the peer receivers in an anonymity system. Hence the message domain $\mathcal{D}_{plain}$ corresponds to the subset of peers that are communication partners of distinct senders in the anonymity system. Letters in the ciphertext can similarly be associated with the pair of sender and receiver anonymity sets $X = (S', R')$ that the attacker can observe from the Mix network at each round. The set of possible sequences of pairs (S', R') therefore represents the ciphertext domain $\mathcal{D}_{cipher}$ in encryption systems.

If the Mix system broadcasts the messages of the senders S' to all possible receivers R of the network instead of only to the real receivers R', then the Mix approach provides perfect relationship anonymity as outlined in [12]. This situation is similar to perfect secrecy; since all possible receivers receive all messages at each round, the attacker gains no knowledge about the relationship between senders and receivers by observing sequences of (S', R').

Nevertheless the overhead of broadcasting is very high, and practical systems do not use this function. Without broadcast it can be shown that an attacker observing sufficiently long sequences of pairs (S', R') can eventually gain enough information to reveal the relationship anonymity of a user. In particular, it has been shown by the disclosure attack[15] that if the attacker observes the pairs (S', R') in which a designated user, whom we call Alice, sends messages then he can unambiguously identify Alice's peer receivers. This assumes that Alice repeatedly communicates with her peer receivers. This feature of anonymity systems is analogous to the ciphertext in encryption systems, which reflects the redundancy of the language. The sequences (S', R') reflect the "redundancy" of Alice's peers. The analogy between the key spaces for the cipher and the Mix system is not obvious at the moment and requires more research, but we can observe that the sequences of the pairs (S', R') are restricted by the number of users in the system, the user's communication partners and the batch size of the Mix.

A further refinement of the disclosure attack is the the hitting-set attack[5]. This attack was proven in [6] to require the least possible number of observations, (S', R'), to unambiguously disclose Alice's peers. Thus, by computing the average length of the sequences (S', R') required to disclose Alice's peers we obtain a measurement that corresponds to Shannon's unicity distance in encryption systems. If the attacker observes less than this number of pairs then the relationship anonymity is information theoretically secure, as Alice's peers cannot be unambiguously identified. Otherwise, information theoretic security cannot be assured. The hitting-set attack assumes the role of a practical brute force attack on the relationship anonymity, and thus as a measurement function for the information theoretic anonymity. This feature enables us to model and analyse anonymity through the hitting-set attack.

4 Related Work

We classify anonymity metrics applicable to the Mix by two main categories: *historyless measurements* and *historied measurements*. These categories are shown in Figure 2. Historyless measurements are characterised by measuring only the anonymity set, without considering information from past anonymity sets, or the potential for those in the future. Examples of these approaches are shown in [8, 13, 9].

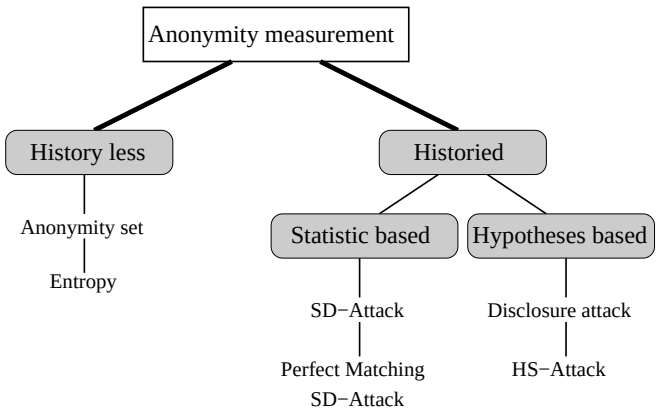

Fig. 2. Categorisation of different anonymity measurement approaches

In contrast are historied measurements that reveal the relationship anonymity between senders and recipients by relating past and future anonymity set relations. The first practical attack to do so was the disclosure attack[15], that showed that measuring the anonymity set alone does not adequately represent the anonymity provided by the system. Following this, the hitting-set attack[5] was introduced, which was proved to require the least number of observations to reveal the relationship anonymity by a passive global attacker[6]. By providing a lower bound for the number of observations required to unambiguously reveal all communication partners of a user, the hitting-set attack provides the basis for an information theoretic measurement of relationship anonymity.

An important strength of this approach is that it is logical and provides provable derivations of possible solutions from the information provided by the anonymity sets. We can therefore logically trace the state of anonymity from the initial state to the state in which all communication partners are revealed. This not only allows us to understand when particular states are reached, but further allows us to model the evolution of anonymity as information is gained. We discuss in this paper possible expansions of analyses and metrics based on this approach.

Other historied measurements rely on analysing the frequency by which a receiver obtains messages[10, 16, 17]. In this approach, the attacker observes the communication round of the network under two different conditions: the first is the observation of those peers that receive a message whenever a particular user sends a message. The second

condition is the observation of the receivers when Alice's does *not* participate. From the observations under these two conditions the attacker derives two communication distributions over the receivers, one with Alice's communication contribution and one without. This second distribution is the *background distribution*. By assuming that the background distribution remains constant over time one can remove the background distribution from the former distribution to produce a set of outstanding peers that may be the sender's peer partners. This approach circumvents the difficulty of analysing the possible solutions, however this simplification is not for free as the approach cannot itself prove that a set of receivers with outstanding distributions is a unique possible solution, nor if it is Alice's peers. This approach therefore cannot answer information theoretic questions.

We now examine in greater detail the development of the traditional metrics for anonymity systems.

4.1 Anonymity Set

One of the first proposed, and simplest, measures for the anonymity afforded to a user in a Mix-style system is the size of the set of users that may have sent a message through the system. As this set becomes larger, the anonymity of the system increases. This technique, the *anonymity set*, was introduced by Chaum in 1988 to analyse his Dining Cryptographer Networks [18] and was used in a number of subsequent analyses of anonymous systems.

The most serious limitation of the anonymity set metric is its failure to express non-uniform probabilities across members of the set of senders. For systems such as the Dining Cryptographer Network, and for cascades of ideal Chaumian Mixes, this limitation is not apparent as sender probabilities are uniform to an observer. Intuitively, a system in which a given user is overwhelmingly likely to have produced a message should be considered less anonymous than one in which all users have equal probability.

Despite its simplicity and lack of applicability in more complex systems, the anonymity set has the advantage of being relatively simple to calculate and can, in general, provide a useful first abstraction of a system's anonymity. As such, it remained the major metric for analyses of anonymity systems for many years.

4.2 Information Theoretic Entropy Metrics

In 2002, as a response to the limitations mentioned above, Danezis and Serjantov [19] and Díaz [20] independently proposed an extension of the anonymity set metric that considered the possibility that members of the anonymity set may have non-uniform possibilities of having sent a particular message. Both approaches measure the information-theoretic entropy of the anonymity set, and thus express the extra information required by an attacker to uniquely identify a given user. Díaz, however, chooses to normalise the entropy with respect to the number of users in the system. In both systems, the anonymity set is thus replace by an *anonymity probability distribution* that allows for a more powerful expression of a system's characteristics in anonymising users.

The core idea behind these information theoretic entropy-based metrics is that the anonymity of a system depends both on the size of the set of possible senders *and the*

uniformity of the probabilities for that set. The more homogenous the set of probabilities, the more information is required on average to describe the system. An attacker must consequently gather a greater number of observations to uniquely identify users.

The approach taken in these two papers has some limitations as a metric for anonymity systems. The foremost of these is that the quantification of anonymity is provided for a particular user, or a particular message, at a given point in the system execution. Whilst this allows for the direct quantification of the anonymity provided by a Mix at a given moment in time, it proves less effective for analysing a system in the abstract case.

More seriously, the reliance on the observations taken at a given point in a given system makes it extremely difficult to compare objectively different anonymity systems. The analysis provided by these methods can only reveal the likelihood of relationships between specific users and messages, and is not immediately suitable for a more general quantification.

These limitations can be overcome, to a certain extent, by a statistical approach. Diaz and Sassaman [21] used a large volume of gathered traffic data from two different Mix-based systems and calculate the maximum and minimum observed entropies over a long time period. This approach overcomes, to some extent, the limitation of the information theoretic entropy-based metrics at the expense of a much greater reliance on simulation.

4.3 Extensions of Information Theoretic Metrics

There have been a number of attempts to extend the applicability of the basic information theoretic entropy-based metrics of anonymity.

As mentioned above, Díaz and Sassaman applied a statistical approach towards the comparison of two implemented Mix systems. This comparison made use of a simulation-based approach, based on large volume of traffic data collected from the existing MixMaster [22] and Reliable anonymous remailer networks. An information theoretic entropy-based measure, broadly similar to the approach of Danezis and Serjantov, was employed to examine the maximum and minimum expected anonymity provided to users under a variety of system settings and traffic conditions.

Wright[23] applied an information theoretic quantification, combined with simulations, to measure the amount of confusion introduce into message ordering by a variety of anonymity mechanisms that included Mixes. In this approach, pairs of messages were injected together into the input message stream of different systems. The distance introduced between these originally adjacent messages in the stream of output messages was then used to calculate a probability distribution for the system over a large number of iterations. By calculating the distribution of the introduced distances, the effectiveness of the systems in “unordering” messages could be calculated and expressed in terms of information theoretic entropy. By considering only the input and output streams of messages, this approach allowed for an the comparison of a number of highly different abstract anonymity systems.

Chatzikokolakis [24] extended the use of entropy for measuring anonymity systems by considering channel capacity, a later development in Shannon’s information theory that measures the information that can be transmitted reliably over a channel.

Chatzikokolakis modelled the entire anonymity systems as an abstract channel that transmits identifying information about users encoded as observations of message flow. The more effective this channel is at transmitting information, the less effective it is as an anonymity system. This approach was applied both to simple Mixes and to Chaum's Dining Cryptographer Networks. A further advantage of this approach is that, from the consideration of channel capacity, the error probability of an attacker in calculating identifying information may easily be expressed according to the number of observations at a given point.

4.4 Conclusion

The development of anonymity metrics has progressed from highly simple quantifications through increasingly complex analyses of the information required to identify users. Later quantifications have sought to express anonymity in a variety of different systems, and to provide more fine-grained quantifications of the anonymity provided. Increasingly, these metrics have attempted to quantify the overall anonymity provided by a system rather than that that may be expected by a given user under given conditions.

There are still limitations, however, in the metrics that have been shown here. Most notably, the traditional anonymity metrics have expressed the amount of extra information required by an attacker to uniquely identify a user, but do not consider the ease with which that information may be obtained. We will now proceed to examine other forms of anonymity metric that seek to resolve the limitations that still exist in these metrics, by considering more directly attacks that result in the unique identification of users.

5 Statistics-Based Measurements

The original statistic based approach was the statistical disclosure attack[10]. Although the name suggests a relationship, this attack only shares the attacker and Mix model with the original disclosure attack. The fundamental idea of the statistical disclosure attack is to count the cumulative messages that each peer in S receives over several rounds. This counting is done for those rounds where Alice participates and those in which Alice does not participate, in order to obtain the background frequency distribution. This results in two probability distribution vectors:$\vec{P}_o$ with Alice's participation and $\vec{P}_u$ without Alice's participation. Each entry in a vector represents the probability that the corresponding receiver is contacted. The difference between these two vectors result in a new probability distribution vector $\vec{P}_{Alice}$ where peers with the most outstanding distributions are considered to be Alice's peers. Details about the computation of $\vec{P}_o$, $\vec{P}_u$ and $\vec{P}_{Alice}$ can be found in [10, 16]. The mean number of observations required such that Alice's peers are outstanding is given by the formula [10]:

$$t_{SD} > \left(lm \left(\sqrt{\frac{m-1}{m^2}} + \sqrt{\frac{N-1}{N^2}(b-1)} \right) \right)^2, \tag{1}$$

where l is factor that determines the confidence of Alice's peers being outstanding. The value $l = 1$ corresponds to an 84% confidence, $l = 2$ to a 97% confidence and $l = 3$ to a 99% confidence.

This approach is applicable only if:

- Alice's peers are significantly more frequently contacted than non-peers,
- the pre-knowledge of the frequency distribution of non-peers is given,
- the frequency distributions of the non-peers do not change noticeably during the attack.

These requirements represent additional restrictions that do not exist in the disclosure attack or the hitting-set attack. In particular, the last point is somewhat unrealistic since it assumes that all senders in the network constantly maintain their communication behaviour. There is also no analysis of what is required for a change in communication frequency to be noticeable. Existing simulations in the literature do not address this problem, and simulate a constant sender behaviour.

This attack does not determine whether the set of most outstanding peers unambiguously represents Alice's peers or what the other solutions may be, as produced by the disclosure attack and hitting-set attack. The results of simulations based on this approach must therefore be considered carefully. Although some work based on this attack does consider the amount of observations for a successful attack, this represents the time taken for the attack to confirm *a priori* knowledge of Alice's peer partners. This is not necessarily related to the genuine identification of Alice's peers, since the criteria for the applicability and the termination of the attack is neither a necessary nor a sufficient condition for the identification of Alice's peer set.

The advantage of the statistical approach is that it is easy to count the required frequencies. Thus, given that all requirements are fulfilled, applying this approach is relatively straight forward.

5.1 Analysis

Figure 3 compares the mean number of observations, t_{SD}, by Equation (1) for $l = 1, 2, 3$ represented by the lines SD_1, SD_2, SD_3 with the mean number of observations required to disclose all of Alice's peers obtained by simulations using the hitting-set attack, represented by the line HS. In this simulation a set of b senders send messages to distinct receivers at each round. Among these b senders is Alice, who has m peer partners. Each of this peers is uniformly contacted by Alice with a probability of $\frac{1}{m}$. The other $(b-1)$ senders choose their receivers uniformly from the set of all possible receivers R with probability $\frac{1}{N}$. This simulation functions under the the assumptions for which Equation (1) is valid. The x-axis shows variation in one of the parameters N, b, m, while the y-axis shows the number of observations.

We can see in these graphs that if Alice's peers are statistically outstanding with at least 97% confidence (SD_3) then, in many cases, Alice's peers can be unambiguously identified before this happens. In contrast to this, if Alice's peers are outstanding with a confidence of 85% (SD_1) then Alice's peers can generally not be identified unambiguously. The relation between the statistical disclosure attack and the identification

of Alice's peer is unclear. Nevertheless, many papers choose $l = 2$ when measuring the mean time to identify Alice's peers because the graph for this value, SD_2 in Figure 3, is close to the mean number of observations required to identify Alice's peers unambiguously.

This result, however, is only a coincidence that holds for some parameters. A counterexample can be seen in Figure 4. We can see that, although Alice's peers are statistically significant with a confidence of 97% after t_{SD_2} observations, it is not possible to unambiguously identify Alice's peers at that time. Thus Alice can still deny that the most significant peers are her peers, as many other possibilities remain. This example underpins that the measurement by the statistical attack does not provide assertions about the necessary or sufficient condition for the identification of Alice's peers.

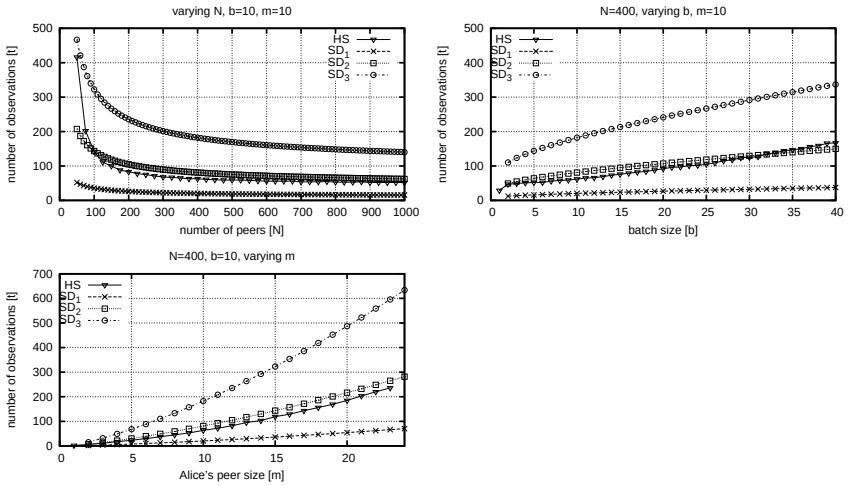

Fig. 3. Comparison of SD_1, SD_2, SD_3 with mean time to disclose all of Alice's peers by HS-attack

Based on the statistical disclosure attack, and under its three main requirements, further statistics-based approaches have been suggested. While the statistical disclosure attack requires a uniform distribution of communication frequencies to peers, this requirement has been relaxed in [16]. A further refinement, the perfect-matching disclosure attack, combines the statistical approach with an anonymity set-based approach[17]. In this work, the statistical disclosure attack is first applied to obtain the distribution vector $\vec{P}_{Alice}$. This vector is then used to weight the sender-receiver relationship probability of the anonymity sets collected in additional rounds. The feedback of $\vec{P}_{Alice}$, combined with the distribution of $\vec{P}_{Alice}$, can be corrected by new observations. This approach can gain more precise results for $\vec{P}_{Alice}$, and is applicable for more varying user behaviour models, and thus relaxes the requirements of the statistical disclosure attack to a certain extent. The drawback of this approach is that it requires significantly more observations

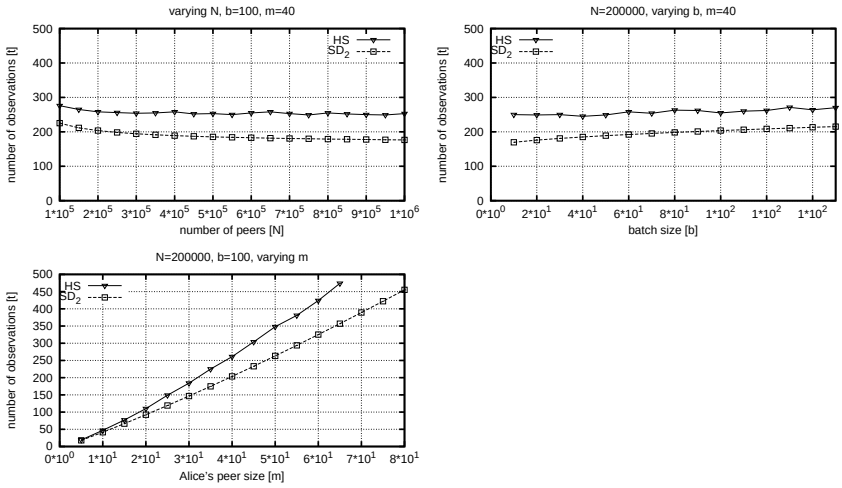

Fig. 4. Comparison of SD_2 with mean time to disclose all of Alice's peers by HS-attack

than the statistical disclosure attack and does not focus on achieving the least number of observations required to identify Alice's peers.

5.2 Information Theoretic Consideration

By focusing on communication frequencies and not on analysing provable interrelations of the considered anonymity sets, statistical approaches leave the answer to the information theoretic question of possible solutions open. Furthermore, these approaches cannot easily measure real-world systems due to their restrictive pre-conditions. Thus, any relation between the statistical significance and the unambiguous identification of Alice's peers would only hold for those communication scenarios in which the pre-conditions are met. An anonymity metric derived from statistical approaches would therefore be of limited validity. Additionally, the amount of information measured by the statistical approaches to reach particular states of anonymity are still not designed to measure precisely the time of disclosing particular information. The number of observations required by the statistical approach may be lower or higher than the amount of observations required to reveal particular information unambiguously.

5.3 Complexity Theoretic Consideration

Complexity theory asks the question: how hard will it be to break a system under the condition that it is information theoretically breakable? Since there is no proven relation between the statistical approaches and the information theoretic consideration, we must consider this aspect under the assumption that a relation does exist. Under this assumption, statistic based approaches will in general not determine how hard it is to disclose a particular anonymity state due to their lack of precision. What we can expect

is that it these approaches may tell us when the disclosure of particular information is easy, by using more information than really necessary.

6 Hypothesis-Based Measures

In our Mix model, it is assumed that Alice keeps communicating with a constant set of communication partners, $\mathcal{H}_A$, during a time period t, where $|\mathcal{H}_A| = m$ it the amount of Alice's peers. Under this condition, the hitting-set attack provides confident and complete knowledge about all reasonable peer sets of Alice[1]. We will call each reasonable peer set $\mathcal{H}$, where $|\mathcal{H}| = m$ represents a ***hypothesis*** and the set of all reasonable peer sets, the ***hypothesis set***. Since the attacker only observes those receiver anonymity sets R' where Alice participates, only those sets which intersect with all the receiver anonymity sets $R'_1, \ldots, R'_t$ observed by the attacker are ***reasonable***. We will call each R' of the pair (S', R') an ***observation*** if $Alice \in S'$. Thus, each hypothesis is the result of the cumulative acquisition of information from all collected observations.

Within the period t, each observation can only give the attacker additional information about Alice's peers, and the hypothesis sets decreases with the increasing number of observations collected by the attacker.

The hypothesis set represents confident knowledge as each hypothesis is a proven hitting set with respect to the attacker's observations. It is complete because it covers all hitting sets. This complete knowledge allows us to measure information theoretically the length of the period t until some particular unambiguous information is revealed. We obtain a measure similar to the unicity distance in encryption systems that shows how long a user can maintain redundant communication behaviour without revealing particular information about his peers. The kind of information that could be revealed is discussed in the following sections.

Note that the most important difference between the hypothesis-based approach and the statistics-based approach is that we deal with confident and complete knowledge rather that statistics. All conclusions derived in this approach are logically verifiable. If we deduce that Alice's peer set can be identified, then this constitutes a proof that no other hypotheses are possible. In contrast to this, the statistical attacks provide only a set of significant frequencies for likely peer partners. Whether these peers are Alice's peers or not, or if there are other consistent hypotheses, cannot easily be determined.

6.1 Information Theoretic Consideration

Figure 5 shows the mean time to disclose all of Alice's peer by a simulated HS-attack, shown on the HS line, together with the mathematically derived time to reduce the hypothesis set to a size below 2, shown on the t_2 line, and the mean time to disclose at least one of Alice's peers, shown on the t_{part}. The model to derive these functions is discussed in the sequel.

[1] Note that statistical approaches are based on the same assumption; we clarify this here to illustrate the similarity to unicity distance.

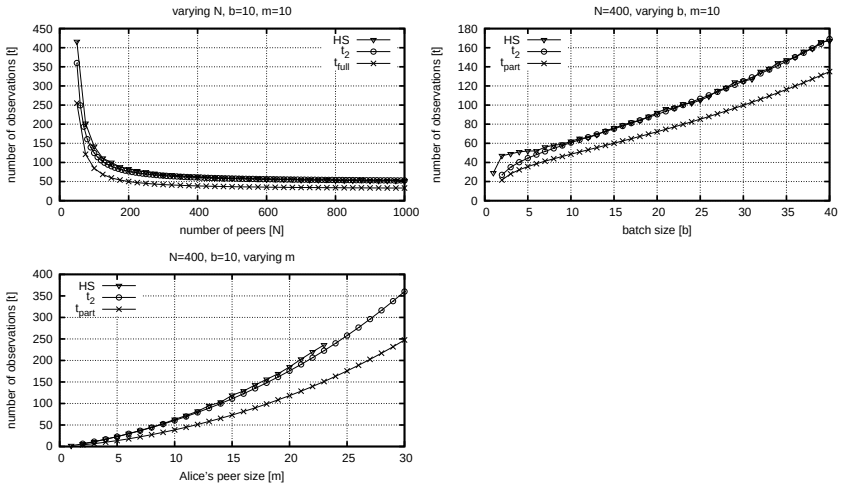

Fig. 5. Comparison: HS-attack simulation, time to reduce hypothesis set below 2 and mean time to disclose at least one Alice's peers

Full Disclosure We define *full disclosure* to be the anonymity state in which Alice's peer set can be unambiguously identified. This state is reflected in the hypothesis set, and is the point at which the hypothesis set consists of exactly one hypothesis. If we measure the mean number of observations, t_{full}, such that the hypothesis set computed by the hitting-set attack contains exactly one hypothesis, then we obtain an analogue to unicity distance for full disclosure. If the attacker observes less than t_{full} observations, or if the period t is less than t_{full}, then Alice's peer set is information theoretically secure. This means that there is at least one other hypothesis, which could also be Alice's peer set. The mean time to disclose Alice's peers is shown in the Figures 3 and 4, labelled HS.

A mathematical lower and an upper bound for t_{full} was first introduced in [6]. Closed formulae for the approximation of t_{full} were provided in [25, 26]. [26] introduced the first mathematic model to describe the content of the hypothesis set with respect to the number of observations collected by the attacker. This enables the study of the evolution of the hypothesis set to different anonymity states. In particular, all anonymity measurements presented in Section 6 can be modelled by the approach of [26]. Nevertheless, this model was developed only for a uniform user communication, and more research is required to extend the model to more realistic distributions.

Hypothesis Set Characteristics The key to analysing the anonymity state of the Mix system is the content of the hypothesis set. Finding characteristics of the hypotheses is therefore crucial to understanding how the hypothesis set is affected if the attack obtains new observations. Apart from increasing our understanding of the change of the hypothesis set over time, these characteristics can also be used to mathematically model the hypothesis set and thus the evolution to distinct anonymity states as shown in

[26]. Under the assumption that Alice has m peer partners, the following characteristics of the hypotheses within a hypothesis set was proven in [26]:

- Each hypotheses is a superset of a minimal hitting set. A *hitting set*, $\mathcal{H}'$, is a set that intersects with all observations of the attacker. It is *minimal* if no proper subset of $\mathcal{H}'$ is a hitting set.
- With increasing observations, all hypotheses become minimal hitting sets.
- The number of hypotheses is strictly bounded by b^m.
- Each hypothesis, $\mathcal{H}$, belongs to exactly one of the structures $\mathfrak{H}_0, \ldots, \mathfrak{H}_m$. $\mathcal{H} \in \mathfrak{H}_i$ if and only if $\mathcal{H}$ contains $(m-i)$ Alice's peers, i.e. $|\mathcal{H} \cap \mathcal{H}_A| = (m-i)$. In particular $\mathfrak{H}_0 = \{\mathcal{H}_A\}$.
- The number of hypotheses of the structure $\mathfrak{H}_i$ is:

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

- If Alice chooses her peers uniformly with probability $\frac{1}{m}$ from $\mathcal{H}_A$, and the other $(b-1)$ senders choose their peers with probability $\frac{1}{N}$ from R at each round, then the probability that a hypotheses of the structure $\mathfrak{H}_i$ is excluded by the next observation is

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

With the main information determined the number of observations t_a, the average size of the hypothesis set is $1 < a < b^m$, starting with an initial hypothesis set of size b^m. For $a = 2$, the value of t_a is the number of observations such that the mean number of hypotheses in the hypothesis set is at most two. This function is shown in Figure 5

$$t_a \leq \frac{m(\ln(b-1) - \ln(a^{1/m} - 1))}{(1 - \frac{m}{N})^{b-1}}$$

Note that the function t_a is not equal to t_{full}. The focus of t_a is the mean size a of the hypothesis set, therefore it is used to compute the number of observations required to reach this mean. In contrast, the focus of t_{full} is the mean number of observations such that the hypothesis set contains exactly one hypotheses, which is $\mathcal{H}_A$. Thus, both t_a and t_{full} are reasonable means to measure the anonymity of the Mix system, but a closed formula for t_{full} has not yet been discovered. Nevertheless, a proven lower bound for t_{full} was introduced in [6].

Partial Disclosure A new direction for measuring the anonymity of a Mix system is the consideration of the point at which at least one of Alice's peer can be unambiguously identified. We call this anonymity state *partial disclosure*. This measurement determines the first unambiguous information revealed by the Mix system. we can also visualise represent this anonymity state with the hypothesis set, it being the point at which all hypotheses have a particular peer in common. Again, we can obtain a unicitiy distance-like anonymity measurement by computing the average number of observations, t_{part}, such that all hypotheses determined by the hitting-set attack contain at least one common peer of Alice.

If less than t_{part} observations are considered by the attack, or if $t < t_{part}$, then each of Alice's peers is information theoretically anonymous; none of Alice's peers can be unambiguously identified. This is a stricter anonymity criteria than full disclosure. Furthermore, the point at which at least one peer of Alice is disclosed is noticeably earlier than the point at which all of Alice's peers can be disclosed. The anonymity provided by Mix systems therefore has to be reconsidered in this context. We further conjecture that measuring anonymity by partial disclosure is more solid than by full disclosure in the sense that dummy traffic, or the change of Alice's communication behaviour, may strongly affect the point of full disclosure while leaving the point of partial disclosure relatively untouched.

The first mathematical model for partial disclosure can be found in [26]. This model uses the hypothesis set characteristics and the uniform communication distribution of Alice and the remaining users described in the last section. That work gives the following formula for the probability of identifying at least one of Alice's peers after at most t observations:

$$f_{id_{any}}(N,b,m,t) = \sum_{s=1}^{m}\left((-1)^{s-1}\binom{m}{s}\frac{\prod_{i=1}^{m}(1-(1-\frac{i}{m}(1-\frac{m}{N})^{b-1})^{t})^{\binom{m}{i}(b-1)^{i}}}{\prod_{i=1}^{m-s}(1-(1-\frac{i}{m}(1-\frac{m}{N})^{b-1})^{t})^{\binom{m-s}{i}(b-1)^{i}}}\right)$$

This combines with the mean time, t_{part}, to disclose at least one of Alice's peers. This function is shown in relation with the simulated mean time for full disclosure and the time to reduce the hypotheses below a size of 2 in Figure 5.

$$t_{part}(N,b,m,t) = \sum_{t=1}^{\infty} t\ \left(f_{id_{any}}(N,b,m,t) - f_{id_{any}}(N,b,m,t-1)\right)$$

A closed formula for t_{part} has not yet been presented, and there is no study showing how partial disclosure is affected by different user communication behaviours. Despite this, the results discussed in this paper show prototypically what can be achieved with the hypothesis set and the identified structures with the current state of research.

Beyond Unambiguity The full disclosure and partial disclosure metrics measure only the amount of observations required to reach full or partial disclosure. We could obtain a more refined anonymity measurement if we considered further properties of the hypothesis set. Assume that t observations remain to reach full or partial disclosure; a more refined information theoretic measurement could, for example, take into account that a hypothesis set in which each hypothesis contains at least $m-1$ of Alice's peers, is less anonymous than a hypothesis set in which each hypothesis contains only one of Alice's peers. In [26] the point at which each hypothesis contained a particular number of observations was measure, but this measurement has not been integrated into the full or partial disclosure measurement.

6.2 Complexity Theoretic Consideration

As of yet, it has not been analysed if there are instances in which the hitting-set attack requires less than exponential run time. Thus, the complexity based anonymity of the Mix system remains open research.

7 Conclusion

We outline in this paper the historical development of anonymity measurement strategies from historyless to historied metrics. Since the anonymity of a system depends on the knowledge of the attacker, which grows over time, it is appropriate to measure the anonymity of a system by considering these historied approaches. We identify two families in the literature that we refer to as statistics-based and hypothesis-based. The main advantage of the statistics-based approaches is their simplicity through reliance on statistics. At the same time this presents a disadvantage because these approaches do not identify unambiguous communication relationships.

As a mathematical consequence, the measurements provided by statistical approaches are inappropriate for measuring the precise point at which communication relationships may be identified. As a practical consequence, a user can deny having contacted the statistically significant peers because the statistical approach cannot provide a logical proof of the relationship between a sender and the statistically identified peers.

In contrast to the statistics-based approaches, hypotheses based measurements relying on the hitting-set attack are precise and complete, and mathematical models and approximations built on this foundation are logically sound. We can consequently gain reasonable and precise understanding of the evolution of the anonymity of a Mix system over time.

The most important information used to analyse and measure the anonymity of a system is the hypothesis set. We outline that structures and bounds of the hypothesis set could be identified, which enables the description of the evolution of the hypothesis set. There are still, however, many open questions regarding hypothesis sets. Analysing hypotheses under the assumption of a uniformly distributed communication can be considered as a proof of concept, but we believe that the model enables more realistic analysis as the identified bounds and structures, apart from the probability P_{inv} to exclude hypotheses, are invariant with the communication distributions of the sender.

Another open question is the detection and classification of instances that may be broken in less than exponential time. Finally, it would be interesting to investigate strategies to increase a user's possible anonymous communication time with as little overhead as possible, and thus to discover strategies to effectively and efficiently "attack the attacker".

References

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

[2] Camenisch, J., Lysyanskaya, A.: An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation. In: Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques (EUROCRYPT '01), London, UK, Springer-Verlag (2001) 93 – 118

[3] Goldschlag, D.M., Reed, M.G., Syverson, P.F.: Hiding Routing Information. In Anderson, R., ed.: Proceedings of Information Hiding: First International Workshop, Springer-Verlag, LNCS 1174 (May 1996) 137–150

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

[5] Kesdogan, D., Pimenidis, L.: The Hitting Set Attack on Anonymity Protocols. In Fridrich, J., ed.: Information Hiding: 6th International Workshop (IH2004). Volume 3200 of LNCS. (May 2004) 326

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

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

[8] Díaz, C., Seys, S., Claessens, J., Preneel, B.: Towards Measuring Anonymity. In: Privacy Enhancing Technologies. Volume 2482 of LNCS. (2002) 54–68

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

[10] Danezis, G.: Statistical Disclosure Attacks: Traffic Confirmation in Open Environments. In Gritzalis, Vimercati, Samarati, Katsikas, eds.: Proceedings of Security and Privacy in the Age of Uncertainty, (SEC2003), Athens, IFIP TC11, Kluwer (May 2003) 421 – 426

[11] 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, Hanscom AFB, MA (August 1978)

[12] Pfitzmann, A.: Diensteintegrierende Kommunikationsnetze mit teilnehmerüberprüfbarem Datenschutz. Volume 234 of Informatik-Fachberichte. (1990)

[13] Serjantov, A., Danezis, G.: Towards an Information Theoretic Metric for Anonymity. In: Privacy Enhancing Technologies. Volume 2482 of LNCS. (Januar 2003) 259–263

[14] Denning, D.E.R.: Cryptography and data security. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1982)

[15] Kesdogan, D., Agrawal, D., Penz, S.: Limits of Anonymity in Open Environments. In Petitcolas, F., ed.: Information Hiding: 5th International Workshop (IH2002). Volume 2578 of LNCS. (October 2002) 53 – 69

[16] Mathewson, N., Dingledine, R.: Practical Traffic Analysis: Extending and Resisting Statistical Disclosure. In: Proceedings of Privacy Enhancing Technologies workshop (PET 2004). Volume 3424 of LNCS. (May 2004) 17–34

[17] Troncoso, C., Gierlichs, B., Preneel, B., Verbauwhede, I.: Perfect matching disclosure attacks. In: PETS '08: Proceedings of the 8th international symposium on Privacy Enhancing Technologies, Springer-Verlag (2008) 2–23

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

[19] Serjantov, A., Danezis, G.: Towards an information theoretic metric for anonymity. In Dingledine, R., Syverson, P., eds.: Proceedings of Privacy Enhancing Technologies Workshop (PET 2002), Springer-Verlag, LNCS 2482 (April 2002)

[20] Diaz, C., Seys, S., Claessens, J., Preneel, B.: Towards measuring anonymity. In Dingledine, R., Syverson, P., eds.: Proceedings of Privacy Enhancing Technologies Workshop (PET 2002), Springer-Verlag, LNCS 2482 (April 2002)
[21] Diaz, C., Sassaman, L., Dewitte, E.: Comparison between two practical mix designs. In: Proceedings of ESORICS 2004. LNCS, France (September 2004)
[22] Möller, U., Cottrell, L., Palfrader, P., Sassaman, L.: Mixmaster Protocol — Version 2. IETF Internet Draft (July 2003)
[23] Wright, J.: Characterising Anonymity Systems. PhD thesis, Department of Computer Science, University of York, York (November 2007)
[24] Chatzikokolakis, K.: Probabilistic and Information-Theoretic Approaches to Anonymity. PhD thesis, Laboratoire d'Informatique (LIX), École Polytechnique, Paris (October 2007)
[25] Kesdogan, D., Pimenidis, L.: The Lower Bound of Attacks on Anonymity Systems – A Unicity Distance Approach. In: First Workshop on Quality of Protection. (September 2005)
[26] Pham, V., Kesdogan, D.: A Combinatorial Approach for an Anonymity Metric. In: Australasian Conference on Information Security and Privacy (ACISP 2009). (2009)