Towards Relations Between the Hitting-Set Attack and the Statistical Disclosure Attack

Dang Vinh Pham, Dogan Kesdogan

Towards Relations between the Hitting-Set Attack and the Statistical Disclosure Attack

Dang Vinh Pham and Dogan Kesdogan

University of Regensburg , Regensburg, Germany

Abstract. The Minimal-Hitting-Set attack (HS-attack) is a well-known, provably optimal exact attack against the anonymity provided by Chaumian Mixes (Threshold-Mixes). This attack allows an attacker to identify the fixed set of communication partners of a given user by observing all messages sent and received by a Chaum Mix. In contrast to this, the Statistical Disclosure attack (SDA) provides a guess of that user's contacts, based on statistical analyses of the observed message exchanges.
We contribute the first closed formula that shows the influence of traffic distributions on the least number of observations of the Mix to complete the HS-attack. This measures when the Mix fails to hide a user's partners, such that the user cannot plausibly deny the identified contacts. It reveals that the HS-attack requires asymptotically less observations to identify a user's partners than the SDA, which guesses them with a given bias. This number of observations is $O(\frac{1}{p})$ for the HS-attack and $O(\frac{1}{p^2})$ for the SDA, where p the probability that the attacked user contacts his least frequent partner.

1 Introduction

Anonymity in communication networks is an essential part of privacy. According to the definition of Pfitzmann et al. [24]: "*Anonymity* is the state of being not identifiable within a set of subjects, the *anonymity set*". Anonymity systems commonly seek to establish anonymity sets. The most influential work in this area is the Chaumian Mix (also known as Threshold-Mix) [7] that forms the basis of many popular services offering anonymity in open and shared networks [29], e.g. the Internet. A Threshold-Mix collects in every *round* a batch of b encrypted messages from distinct senders, who all contribute the same number of messages[1] of identical size. It changes the appearance and time characteristics of the messages in the output batch to provide unlinkability between its input and output messages. Therefore, the senders and recipients that use the Mix in a round form the sender- and recipient-anonymity set in that round.

This work investigates the fundamental limit of anonymity provided by the anonymity sets established by the Threshold-Mix with respect to a *global passive attacker*[2]. Analogous to the fundamental work of Shannon's unicity distance [30], we focus on determining the number of observations of Mix rounds required to disclose a profile of

[1] Otherwise, it would be trivial to identify a pair of sender and a recipient by the number of their exchanged messages in a round.

[2] This attacker can observe any link in the network and can thus observe the anonymity sets.

an arbitrary user (say Alice) and thus to break the anonymity system. We consider the case that Alice's profile determines a static set of friends that are repeatedly contacted by Alice. It is motivated by the observation that human relationships tend to be persistent and by the fact that anonymity should also be provided in this case.

The immanent information leaked by the Mix to a global passive attacker is the observed set of senders and recipients using the Mix in a round. Traffic analysis attacks can learn Alice's profile by accumulating this information, although the Mix provides unlinkability between the input and output messages in a single round. We distinguish between two categories: *combinatorial* attacks [4, 2, 17, 16, 26, 27] and *statistical* attacks [8, 9, 20, 11, 10, 31, 23, 22]. Combinatorial attacks are basically concerned with the disclosure of exact information about Alice's profile that is consistent to the observations of the anonymity system. In contrast to that, statistical attacks are concerned with classifying whether a recipient is likely Alice's friend, or not. Their main advantage is the computational efficiency. However, combinatorial attacks (e.g., the HS-attack) can also be computational efficient [27] for non-trivial cases. The classification of recipients by statistical attacks can lead to a profile that deviates from Alice's profile, e.g., due to false-positive errors, which classify recipients as friends that are not Alice's friends, or due to omitting friends.

We consider in this work the *Minimal-Hitting-Set* attack (HS-attack) [17, 27], a combinatorial attack that provably requires the least number of rounds to uniquely identify Alice's set of friends [16]. Therefore it determines the fundamental limit of anonymity provided by the Threshold-Mix. This number of rounds is dependent on the traffic distribution of the users and on the parameters of the anonymity system. We contribute a closed formula that estimates the mean of this number with respect to arbitrary distributions of Alice's communication and the parameters of the Threshold-Mix. This complements past works that could only model uniform traffic distributions [18, 16, 26], which are less realistic. Therefore, we are to the best of our knowledge the first to provide such an analytical estimate. Our estimate proves that the number of rounds to uniquely identify Alice's set of friends by the HS-attack is $O(\frac{1}{p})$, while it is $O(\frac{1}{p^2})$ to classify all friends with some error rate by the SDA. The probability $0 < p < 1$ denotes the least probability in the distribution of Alice's traffic to her friends.

Although this work mainly addresses the anonymity of the Threshold-Mix, it might be generalisable to analyse the anonymity of other Mix variants like the *Pool-Mix* [29] that models Mixmaster. There are initial works towards this direction [25, Chap. 5.2] that extends the HS-attack for the Pool-Mix and identifies some conditions for the disclosure of Alice's set of friends in that Mix.

Our analyses refer to high-latency Mix systems, as they seek to protect against global passive attackers. In contrast to these, low-latency systems like Tor [13] and JAP [3, 19] (as applied in practice) do not try to withstand a global attacker in their design.

1.1 Related Works

Our work is concerned with *passive traffic analysis* attacks [29]. These rely solely on external traffic observations of an anonymity system.

The idea of combinatorial traffic analyses was first discussed by Raymond [29] who also sketched the "intersection attack". Later two implementations of combinatorial approaches have been suggested in parallel, the *Intersection attack* [4] and the *Disclosure attack* [2]. The first approach identifies the recipient of a targeted sender for the case that this sender repeatedly contacts a recipient from a singleton [4]. In contrast to this, the Disclosure attack uncovers an arbitrary large set of repeated contacts of the targeted sender, which is thus more general than the Intersection attack. These were followed by the HS-attack [17], that unambiguously identifies a user's communication partner set with a provably minimal number of observations [16].[3] The limitation to all these attacks are that they require the solution of an NP-complete problem [14] to succeed, placing a high computational burden on the attacker. However, the most recent HS-attack that uses the ExactHS algorithm [28, 27] achieves a mean polynomial computational complexity for many non-trivial Mix configurations as proved in [27]. Due to the optimal nature of the HS-attacks, the observations required to conduct them provide a measurement for the anonymity provided by Mix system. Estimates of this number were suggested in [18, 16, 26, 21] for a simple model of uniformly distributed communication traffic.

Statistical attacks identify users through statistical patterns in traffic data. These attacks, introduced by the *Statistical disclosure* attack (SDA) [8, 9, 20, 11, 10], and subsequently improved by the *Perfect-matching disclosure attack* (PMDA) [31] and the *Bayesian-interference* [12], achieve significant increases in computational efficiency by relaxing the requirement for absolute correctness and allowing misclassification or omission of actors. The *Least square approach*[23] attempts to analytically analyse the deviation between a user's profile and the classification provided by it for the Threshold-Mix. Provided the same Threshold-Mix model as in SDA [8] (that is often used in combinatorial analyses, as well as in this work in Section 2) this approach is identical to the SDA. A succeeding extension [22] of this approach considers analogous analyses for the Pool-Mix model.

1.2 Structure

We introduce a simple model for the Threshold-Mix and the attacker, as well as the scheme of the HS-attack in Section 2. Section 3 estimates the mean least number of rounds to uniquely identify Alice's set of friends by a closed formula, based on this model. It compares this estimate with the number of rounds required by the SDA mathematically which shows that the SDA requires asymptotically more observations. Our analyses are confirmed and illustrated by simulations and mathematical evaluations in Section 4. Section 5 finally concludes the work and suggests future works. The proofs of all claims are provided in Appendixes A.

[3] The intersection attack is identical to the special case of the HS-attack, where a targeted sender has exactly one recipient.

2 Mix and Attacker Model

We consider the Mix system as a black box that outputs information that is visible to the attacker (i.e. the sender-anonymity sets and recipient sets), as illustrated in Fig. 1. It represents a generalised and simplified model of practical real-world Threshold-Mixes.

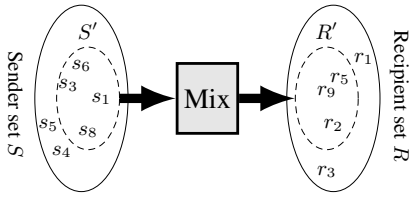

Fig. 1. Mix model.

The Mix is abstractly described as follow:

- A communication system consists of a set of senders, S, a set of all recipients, R, and a Mix node as shown in Fig. 1. S and R represent all users with the ability to send or receive messages in the system[4]. If a sender $s \in S$ communicates with a recipient $r \in R$, then we say that r *is a recipient of* s.
- In each communication round, a subset $S' \subseteq S$ of all senders each send precisely one message to their recipients. Let $R' \subseteq R$ be the set of intended recipients.
- We call S' the *sender-anonymity set*, which is the set of all senders that may have sent a given message in a round. The *recipient set* R' is the set of all recipients that have received a message in a round.
- We label the size of the sender-anonymity set, $|S'|$, as b which is also called the *batch size*.
- The size of the *recipient set*, $|R'|$, is less than or equal to b, as each sender sends exactly one message per round, but several senders may communicate with the same recipient. The size of the set of all recipients is $|R| = u$.

2.1 Attacker Model

We consider a *global passive* attacker that observes the traffic on all links between the user and the Mix in the network. Therefore, he can observe all sending and receiving events in the Mix system, so that the pairs of sender anonymity set and recipient set (S', R') of every round is known to the attacker.

The goal of the attacker is to compute, from a set of traffic observations, all possible sets of friends of a target sender $Alice \in S$. These possibilities form *hypotheses* for the true set of Alice's set of friends, ${}_A\mathcal{H}$, which is assumed to be a fixed set of size $m = |{}_A\mathcal{H}|$. We call a recipient $r \in {}_A\mathcal{H}$ a *friend*; a recipient that does not communicate with Alice, $r \in R \setminus {}_A\mathcal{H}$, is called a *non-friend* and r is simply called a *recipient* if no distinction is required. To clarify that a variable $r \in R$ refers to a friend, it is also denoted by a, whereas it is denoted by n, if it refers to a non-friend.

The attacker focuses on revealing Alice's set of friends by observing only those pairs (S', R'), where Alice participates as a sender. Under this condition we refer to

[4] This definition allows for cases of $S \neq R$, as well as $S = R$, i.e. the sender and recipient set might be distinct or identical.

the corresponding recipient set R' as an *observation*, $\mathcal{O}$. The set of all observations collected during t communication rounds is referred to as the *observation set* $\mathcal{OS} = \{\mathcal{O}_1, \ldots, \mathcal{O}_t\}$.

2.2 Hitting-Set Attack

Alice's possible set of friends can be specified by computing all hitting-sets of size m with respect to the observation set $\mathcal{OS}$ collected by the attacker. A *hitting-set* is a set that intersects with all observations[5] in $\mathcal{OS}$. A hitting-set is a *minimal-hitting-set* if no proper subset of it is a hitting-set. We call a hitting-set $\mathcal{H}$ a *unique minimum-hitting-set*[6], if all hitting-sets $\mathcal{H}' \neq \mathcal{H}$ in $\mathcal{OS}$ fulfil the condition $|\mathcal{H}| < |\mathcal{H}'|$.

By collecting sufficiently many observations, until $\mathcal{OS}$ contains a unique minimum-hitting-set, the attacker can unambiguously identify Alice's set of friends ${}_A\mathcal{H}$. The intuition behind this attack is that at least one of Alice's friends in ${}_A\mathcal{H}$ appears in each observation (due to the definition of observations), while this does not hold for any set $\mathcal{H} \not\supseteq {}_A\mathcal{H}$. Therefore, if there are sufficiently many observations, then ${}_A\mathcal{H}$ becomes a unique minimum-hitting-set. This attack is known as the *Minimal-Hitting-Set attack* (HS-attack)[17]. We refer in the remaining paper to its most recent version that uses the ExactHS algorithm to compute the minimal-hitting-sets [27]. The HS-attack repeats aggregating new observations and computing all minimal-hitting-sets of a given size m' in the aggregated observation set $\mathcal{OS}$. It is successively applied for $m' = 1, \ldots, m$. If m' underestimates m, then there will be no hitting-set of size m' after a sufficient number of observation. This can be detected by the HS-attack to consider a larger value of m' in the HS-attack, until $m' = m$ and ${}_A\mathcal{H}$ becomes a unique minimum-hitting-set. As proved in [16], the HS-attack requires the least number of observations to uniquely identify Alice's set of friends with respect to the Threshold-Mix.

Attack Scheme In our Mix and attacker model, the effort of identifying Alice's set of friends is dependent on the *Mix parameters* (u, b, m) and the distribution of the *cover traffic* and of *Alice's traffic*. The cover traffic is induced by the communication of senders other than Alice to the recipients in the observations. We use the term *Mix configuration* to refer to a combination of Mix parameters and these traffic distributions. The basic scheme underlying the analysis of the HS-attack is illustrated in Fig. 2.

Alice's traffic distribution is modelled by the probability mass function $P_A(a)$ for $a \in {}_A\mathcal{H}$, where $\sum_{a \in {}_A\mathcal{H}} P_A(a) = 1$. The cover traffic distribution is indirectly modelled by the probability function $P_N(r)$, which is the probability that any $b-1$ senders (other than Alice) of a batch contact the recipient $r \in R$ in an observation.

3 Mean Number of Observations for Unique Identification

It was proved in [16] that the $2\times$-exclusivity of Alice's set of friends is a necessary condition for the unique identification of Alice's set of friends. The number of observa-

[5] Due to the definition of observations, ${}_A\mathcal{H} \cap \mathcal{O} \neq \emptyset$ for all $\mathcal{O} \in \mathcal{OS}$, therefore ${}_A\mathcal{H}$ is a hitting-set in $\mathcal{OS}$.

[6] Every unique minimum-hitting-set is a minimal-hitting-set, but not reversely.

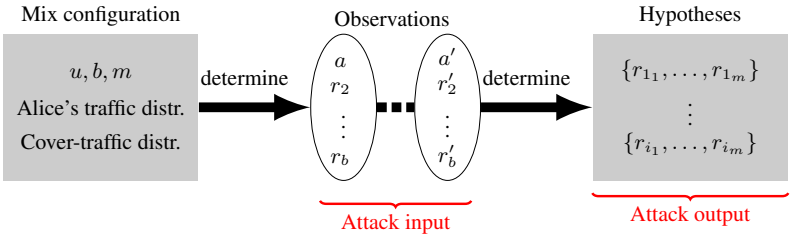

Fig. 2. Analysis scheme: Variables a, r represent arbitrary friend $a \in {}_A\mathcal{H}$ and recipient $r \in R$.

tions aggregated by the attacker, until the $2\times$-exclusivity condition is fulfilled provides a close estimate of the least number of observations to uniquely identify Alice's set of friends, as evaluated in [16].

We contribute a closed formula that estimates the expected least number of observations to fulfil $k\times$-exclusivity, which is for the general case of $k \in \mathbb{N}$. As defined in [16], a friend $a \in {}_A\mathcal{H}$ is *exclusive*, if there is an observation $\mathcal{O}$ that contains only a as an Alice's friend. This means $\mathcal{O} \cap {}_A\mathcal{H} = \{a\}$ and we call $\mathcal{O}$ the *observations that contains a exclusively*. A friend $a \in {}_A\mathcal{H}$ is *$k\times$-exclusive*, if it appears at least k times exclusively in observations, or at least one time alone in an observation (i.e. there is an observation $\mathcal{O}' = \{a\}$). The *$k\times$-exclusivity* is fulfilled, if all Alice's friends are $k\times$-exclusive.

3.1 Mean Number of Observations for $k\times$-Exclusivity

We estimate the mean of the least number of observations $E(T_{k\times e})$ for $k\times$-exclusivity by decomposing this mean in two sub means and estimating those sub means. These are the estimates of the following means[7]:

- The mean least number of observations $E(T_{k\times})$, until Alice contacts all her friends at least k times. This is regardless whether the observations are exclusive, or not.
- The maximum of the mean least number of times $E(T_{e,a})$ Alice has to contact a given friend $a \in {}_A\mathcal{H}$, until it is exclusive, with respect to all Alice's friends $a \in {}_A\mathcal{H}$. For each given friend $a' \in {}_A\mathcal{H}$, this mean only accounts those observations, where Alice contacts a', and the maximum of that mean is $\max_{a' \in {}_A\mathcal{H}} E(T_{e,a'})$.

The variables $T_{k\times e}$, $T_{k\times}$ and $T_{e,a}$ are random variables for: the least number of observations to fulfil $k\times$-exclusivity, the least number of observations until Alice contacts all friends at least k times and the least number of times Alice has to contact a friend a, until it is exclusive. We define $E(T_e) = \max_{a' \in {}_A\mathcal{H}} E(T_{e,a'})$ and set for $a = \operatorname{argmax}_{a' \in {}_A\mathcal{H}} E(T_{e,a'})$, the equality $T_e = T_{e,a}$.

Note that the value of $T_{k\times}$ is dependent on Alice's traffic to her friends, but is independent of the traffic of other senders. In contrast to that, the value of $T_{e,a}$ depends on whether any sender other than Alice contacts any friend in ${}_A\mathcal{H} \setminus \{a\}$ in observations where Alice contacts a. This is dependent on the cover-traffic, but is independent of Alice's traffic. Therefore, $T_{k\times}$ and T_e are statistically independent.

[7] The composition of theses estimates in Claim 1 provide an estimate of $E(T_{k\times e})$.

Claim 1 *Let $E(T_{e,a})$ be the mean least number of times Alice has to contact a friend $a \in {}_A\mathcal{H}$,* [8]*until a is exclusive and $E(T_e) = \max_{a \in {}_A\mathcal{H}} E(T_{e,a})$. Let $E(T_{k\times})$ be the mean least number of observations until Alice contacts all her friends at least k times*[9]*, for $k \in \mathbb{N}$. The mean least number of observations until all Alice's friends are $k\times$-exclusive is estimated by:*

$$\begin{aligned} E(T_{k\times e}) &\leq E(T_{k\times})E(T_e) \\ &\approx \left(\frac{1}{p}(\ln m + \gamma) + (k-1)\frac{1}{p}\ln\ln m\right)\left(\frac{u-(m-1)}{u}\right)^{1-b} \quad , \end{aligned} \tag{1}$$

where $p = \min_{a \in {}_A\mathcal{H}} P_A(a)$ and $\gamma \approx 0,57721$ is the Euler-Mascheroni constant.

We conclude by (1) that the $2\times$-exclusivity of all Alice's friends requires on average $\left(\frac{1}{p}(\ln m + \gamma) + \frac{1}{p}\ln\ln m\right)\left(\frac{u-(m-1)}{u}\right)^{1-b}$ observations. The proof of this claim can be found in Appendix A.

3.2 Relation to Statistical Disclosure Attack

While the HS-attack aims at exact identification of friends; statistical attacks, as introduced by the SDA, cf. [9], aim at correct classification of friends with some probabilities. Although these two approaches are orthogonal, we can now analytically compare the number of observations required by these attacks by (1).

The SDA [9] considers the classification of each friend as a signal to noise problem. It virtually interprets Alice's traffic volume to a friend $a \in {}_A\mathcal{H}$ as a signal and the cumulative traffic volume of other senders to any recipient $r \in R$ as a noise. Let t be the number of observations and p be the probability that Alice contacts a in an observation, then the mean signal to a is pt with the variance $p(1-p)t$. To simplify the maths it is assumed that every non Alice sender contacts a recipient uniformly distributed, so that $\frac{1}{u}$ is the probability that r is contacted by a single non Alice sender. As there are $b-1$ non Alice senders in a batch, the mean noise to a recipient r after t observations is $P_N(r) = \frac{1}{u}(b-1)t$, with the variance $\frac{1}{u}(1-\frac{1}{u})(b-1)t$.

The SDA classifies a friend a better than a random guess, if the mean signal to a is higher than the sum of the standard deviation of the signal and of the noise to a [9]. This is a necessary condition to distinguish the signal from the noise to the same recipient. The least number of observations, such that this condition is fulfilled with a probability determined by a confidence parameter l is, cf. [9],

$$\frac{1}{p^2}l^2\left[\sqrt{\frac{u-1}{u^2}(b-1)} + \sqrt{\frac{u-1}{u^2}(b-1) + p^2(\frac{1}{p}-1)}\right]^2 \quad . \tag{2}$$

Setting $l = 2$, $l = 3$ in (2) leads to a classification with a true-positive rate of 95%, respectively 99%. Let us set $p = \min_{a \in {}_A\mathcal{H}} P_A(a)$, as the recipient which is least frequently contacted by Alice dominates the number of observations to classify all friends. In the case that Alice's traffic is uniformly distributed, $p = \frac{1}{m}$ as assumed in [9].

[8] This only refers to observations, in that Alice contacts a, that is $\mathcal{OS}_A[a]$.

[9] This is regardless whether the observations are exclusive, or not.

We can now compare (2) with (1) (for $k = 2$) with respect to the probability p by fixing all other parameters u, b, m, l; they are identical in both equations. This reveals that the SDA requires $O(\frac{1}{p^2})$ observations to classify all Alice's friends while the HS-attack only requires $O(\frac{1}{p})$ observations to uniquely identify all Alice's friends. This relation between the HS-attack and the SDA is visualised for some examples in Section 4.

4 Evaluation

This section illustrates the closeness of the estimate of the least number of observations to identify Alice's friends and compares this with the number of observations required by the SDA.

The first task applies the $2\times$-exclusivity evaluation an the HS-attack on simulated random observations of a Threshold-Mix. These empirically measure the least number of observations for the $2\times$-exclusivity and the identification of all friends. We use them to illustrate the closeness of the corresponding mathematical estimate by (1).

The second task compares the estimated mean number of observations required by the HS-attack and the SDA for some Mix configurations considered in the simulations. This illustrates that SDA requires asymptotically more observations than the HS-attack.

The traffic distributions that we use to model Alice's traffic and the cover traffic in all simulative and mathematical evaluations are described next.

- Alice contacts in each observation a friend that is randomly drawn from a *Zipf*(m, α) distribution of ${}_A\mathcal{H}$. The probability that she contacts her i-th most frequent contact is $P_A(a_i) = P_z^{m,\alpha}(i) = \frac{i^{-\alpha}}{\sum_{l=1}^{m} l^{-\alpha}}$, where $P_z^{m,\alpha}(i)$ is the probability mass function of the Zipf(m, α) distribution. Note that ${}_A\mathcal{H}$ is uniformly distributed if $\alpha = 0$.
- The remaining $b - 1$ recipients of the cover traffic in an observation are drawn uniformly from the set of $|R| = u$ possible recipients. This means that for all $r \in R$, the probability that any of the $b - 1$ senders other than Alice contacts r in an observation is $P_N = 1 - (\frac{u-1}{u})^{b-1}$.

Alice's traffic is modelled by a Zipf distribution, as it is known to closely model e-mail and internet traffic [1, 6, 15]. An example of this distribution is illustrated in Fig. 3 for distinct values of α, provide that Alice has $m = 23$ friends. The cover-traffic is for simplicity modelled by a uniform distribution that represents a bound of the real distribution. Note that an observation contains the recipients of senders who randomly communicate in the same round as Alice and is therefore a random variable. The distribution of this random variable and thus the number of observations to identify Alice's friends is dependent on the overall distribution of the cover-traffic and of Alice's traffic, regardless of differences in the communication distribution of the individual senders of the cover-traffic.

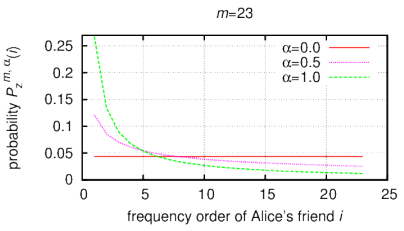

Fig. 3. Zipf(m, α) distribution of Alice's friends, for $m = 23$.

Therefore we assume that all non-Alice senders behave the same to simplify the maths and the simulation.

The HS-attack is *successful* (or *succeeds*) if it uniquely identifies Alice's set of friends ${}_A\mathcal{H}$. For a given Mix configuration, the simulation generates new random observations until the HS-attack is successful and we call this an *experiment*. The average number of observations required by an attack is therefore the mean of the number of observations of all successful attacks (i.e. of all experiments with the same Mix configuration). Note that the results of these experiments, i.e., the number of observations to succeed the HS-attacks, are identically distributed independent random variables with unknown positive mean μ and standard deviation σ. By the law of large numbers, the empirical mean of the experiments' results approaches μ, while its standard deviation approaches 0, for large number of experiments[10]. To ensure that our results are statistically significant, experiments with the same Mix configuration are repeated until 95% of the results fall within an interval of 5% around the empirically observed mean. Every experiment is repeated at least 300 times and no experiment is dropped. We observed that most of our experiments require no more than 300 repeats to fulfil the statistical significance condition and therefore chose this number as a lower threshold. It is necessary to force a sufficiently large least number of repetitions to avoid cases like, e.g., after running two experiments, both results are within 5% around the empirical mean, which would be too few to represent a reliable measure.

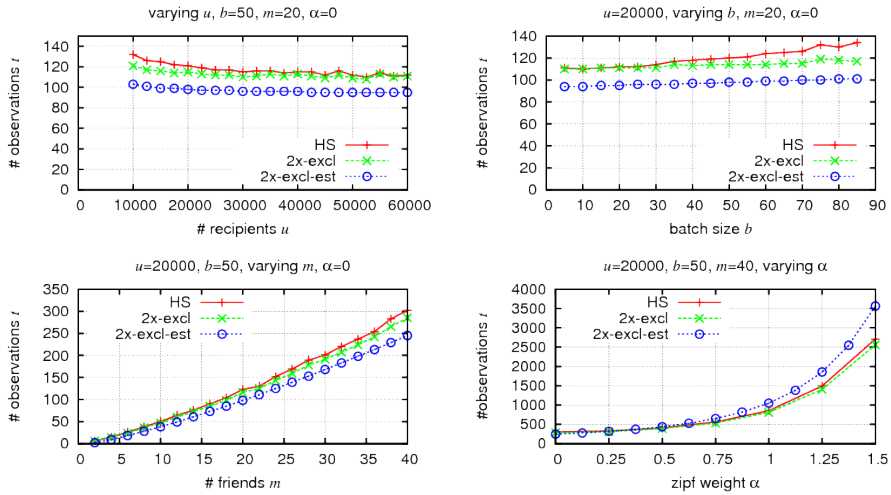

Fig. 4. Mean number of observations: to succeed HS-attack (HS) and to fulfil 2×-exclusivity (2x-excl) versus estimated mean for 2×-exclusivity (2x-excl-est).

Number of Observations Required by HS-attack Fig. 4 visualises the empirical mean number of observations to succeed the HS-attack, labelled (HS) and to fulfil 2×-exclusivity labelled (2x-excl), obtained from simulations. These are compared with the estimate (1) of the mean of the least number of observations for 2×-exclusivity, labelled

[10] This law applies regardless of the magnitude of the variation of the results of single experiments.

(2x-excl-est), which is: $E(T_{2\times e}) \approx \frac{1}{p}\left((\ln m + \gamma) + \ln\ln m\right)\left(1 - \frac{(m-1)}{u}\right)^{1-b}$. Since Alice's traffic is Zipf(m, α) distributed, we get $p = \min_{a \in_A \mathcal{H}} P_A(a) = P_z^{m,\alpha}(m)$.

The plots provide these comparisons for distinct Mix configurations that are modelled by the parameters u, b, m, α. The y-axis always shows the mean number of observations, while the x-axis vary one of the parameters u, b, m, α. We can observe that the estimate (1) provides reasonable approximations, even for the cases that Alice's traffic is non-uniformly distributed (i.e. $\alpha > 0$). According to [1, 6, 15], the value of $\alpha \approx 1$ typically models a user's Zipf(m, α) distributed traffic in the Internet. Due to a lack of experiences with running high-latency anonymity systems in a large user base, we have no authentic empirical values for the parameters u, b. Therefore, we choose parameter ranges that would be reasonable for JAP. JAP was designed to be close to the Chaum Mix, so that it contains batch mixing capabilities [19]. However, collecting messages for a batch increases the latency that is yet not accepted by many JAP users [19], so that this function is disabled in favour of low-latency. Therefore we refer to JAP as a low-latency system. The total number of users that repeatedly use the Dresden-Dresden JAP cascade is about 50000 [19] in 2009, therefore we consider u in the range up to 60000. In every minute, the cascade relays on average 17000 HTTP messages [19], which are 283 messages per second. JAP allows users to send multiple parallel messages, so that the number of messages per second would be lower, if every user is only allowed to send one message in a Mix round to prevent linking a communication by packet counting, as in the Chaum Mix [7]. Simulating batch sizes of up to 85 thus appears to be of reasonable order.

Number of Observations Required by HS-Attack vs. SDA We illustrate that the SDA requires a number of observations that is by the factor of $O(\frac{1}{p})$ higher than those required by the HS-attack, where p is the least probability in the distribution of Alice's friends.

Table 1. Estimated number of required observations: HS-attack (2x-excl-est) versus SDA with 95% true-positive classification (SDA$_{95\%}$).

$u = 400$, $b = 10$, $m = 23$, varying α

α	p	2x-excl-est	SDA$_{95\%}$
0.0	0.0435	186	343
0.5	0.0253	319	840
1.0	0.0116	693	3282
1.5	0.0041	1960	23036

$p = \min_{a \in_A \mathcal{H}} P_A(a) = P_z^{23,\alpha}(23)$ in Zipf$(23, \alpha)$ distribution

$u = 20000$, $b = 50$, $m = 40$, varying α

α	p	2x-excl-est	SDA$_{95\%}$
0.0	0.0250	245	291
0.5	0.0140	437	637
1.0	0.0058	1047	2301
1.5	0.0017	3564	17586

$p = \min_{a \in_A \mathcal{H}} P_A(a) = P_z^{40,\alpha}(40)$ in Zipf$(40, \alpha)$ distribution

Table 1 provides evaluations for the Mix parameters $(u = 400, b = 10, m = 23)$, respectively $(u = 20000, b = 50, m = 40)$ and Zipf(m, α) distributed Alice's traffic. The cover-traffic is uniformly distributed. The tables list the estimated number of observations to succeed HS-attack based on (1) labelled by (2x-excl-est) and to classify Alice's friends with a true-positive rate of 95% by the SDA based on (2) (for $l = 2$) labelled by (SDA$_{95\%}$). We observe that the number of observations required by the

SDA increasingly exceeds that required by the HS-attack for increasing value of α, as p decreases with increasing α.

Note that (2) solely considers the true-positive rate of the SDA; the classification of a given friend as a friend with a certain rate (e.g. 95% in Table 1). However, the false-positive rate can be larger. When SDA terminates, there is thus some number of non-friends that are classified as friends, whereas there is a unique identification of Alice's set of friends, when HS-attack terminates.

5 Conclusion

Anonymous communication systems seek to embed senders and recipients in anonymity sets to hide their communication relations. We measure in this work the anonymity provided by the anonymity sets constructed by the Threshold-Mix to analyse its limit of achievable protection. This limit is determined by the least number of observations of the Mix rounds, until Alice's set of friends can be exactly identified, so that the protection provided by the Mix is repealed. Alice's set of friends can be exactly identified with the least number of observations by the HS-attack [16].

We contribute by (1) (for $k = 2$) the first closed formula that estimates the mean least number of observations to uniquely identify Alice's set of friends for arbitrary distribution of her traffic. It reveals that this number is $O(\frac{1}{p})$, whereas the SDA requires $O(\frac{1}{p^2})$ observations to classify Alice's friends with some error. The variable $p = \min_{a \in {}_A\mathcal{H}} P_A(a)$ denotes the least probability in the distribution of Alice's communication to her friends [11]. This implies that the difference between these two number of observations is for more realistic (non-uniform) distribution of Alice's friends notably higher than for the uniform distribution considered in past mathematical analyses [16, 8]. Section 4 experimentally confirms this difference for some zipf distributed communication of Alice which is known to model real e-mail traffic distribution [1, 6, 15]. Alice's set of friends can thus be exactly identified with a number of observations that is asymptotically less than required by the inexact SDA. This exact identification can be even computational feasible for non-trivial cases by using the HS-attack [27].

Our analysis shows that the mean least number of observations for the exact identification is lowest, if Alice's friends are uniformly distributed. Past works [18, 16, 21, 26, 27] that measure the anonymity of the threshold Mix by the time of exact identification assume for simplicity that uniform distribution. Therefore, we can now confirm that those works address a lower bound of the anonymity of Alice's set of friends.

This work explores the least number of rounds of the Threshold-Mix, such that the attacker's uncertainty about Alice's set of friends becomes 0, as a measure of anonymity. Future works might generalise this approach to quantify the attacker's uncertainty about the possible Alice's set of friends with respect to the number of observed rounds of some Mix. This would enable a more fine granular anonymity measure beyond the time of exact anonymity disclosure (i.e., 0 uncertainty), so that we can also analyse the

[11] If Alice's friends are uniformly distributed, then $p = P_A(a) = \frac{1}{m}$ for all $a \in {}_A\mathcal{H}$, otherwise $p < \frac{1}{m}$.

anonymity provided by other Mix variants like the Pool-Mixe. Pool-Mixes [29] operate like the Threshold-Mix, but they can delay the relay of a random selection of messages in the Mix, as implemented in Mixmaster. Therefore, an attacker might observe a recipient set that misses the user that Alice contacts in the observed round. Such observations induce additional uncertainty about the possible Alice's set of friends in the generalised anonymity quantification approach so that the attacker's uncertainty might remain above 0.

References

[1] Adamic, L.A., Huberman, B.A.: Zipf's Law and the Internet. Glottometrics 3, 143 – 150 (2002)

[2] Agrawal, D., Kesdogan, D., Penz, S.: Probabilistic Treatment of MIXes to Hamper Traffic Analysis. IEEE Symposium on Security and Privacy 0, 16 (2003)

[3] Berthold, O., Federrath, H., Köpsell, S.: Web Mixes: A System for Anonymous and Unobservable Internet Access. In: Designing Privacy Enhancing Technologies, LNCS, vol. 2009, pp. 115 – 129. Springer (2001)

[4] Berthold, O., Langos, H.: Dummy Traffic against Long Term Intersection Attacks. In: Privacy Enhancing Technologies, LNCS, vol. 2482, pp. 110 – 128. Springer (2003)

[5] Brayton, R.K.: On the Asymptotic Behavior of the Number of Trials Necessary to Complete a Set with Random Selection. Journal of Mathematical Analysis and Applications 7(1), 31 – 61 (1963)

[6] Breslau, L., Cao, P., Fan, L., Phillips, G., Shenker, S.: Web Caching and Zipf-like Distributions: Evidence and Implications. In: Proceedings of Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM '99. vol. 1, pp. 126 – 134. IEEE (1999)

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

[8] Danezis, G.: Statistical Disclosure Attacks: Traffic Confirmation in Open Environments. In: Proceedings of Security and Privacy in the Age of Uncertainty. pp. 421 – 426 (2003)

[9] Danezis, G.: Better Anonymous Communications. Ph.D. thesis, University of Cambridge (2004)

[10] Danezis, G., Diaz, C., Troncoso, C.: Two-sided Statistical Disclosure Attack. In: In Privacy Enhancing Technologies, LNCS, vol. 4776, pp. 30 – 44. Springer (2007)

[11] Danezis, G., Serjantov, A.: Statistical Disclosure or Intersection Attacks on Anonymity Systems. In: Information Hiding, LNCS, vol. 3200, pp. 293 – 308. Springer (2005)

[12] Danezis, G., Troncoso, C.: Vida: How to Use Bayesian Inference to De-anonymize Persistent Communications. In: Privacy Enhancing Technologies, LNCS, vol. 5672, pp. 56 – 72. Springer (2009)

[13] Dingledine, R., Mathewson, N., Syverson, P.: Tor: The Second-Generation Onion Router. In: Proceedings of the 13th USENIX Security Symposium. pp. 303 – 320. USENIX (2004)

[14] Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. Freeman (1990)

[15] Glassman, S.: A Caching Relay for the World Wide Web. Computer Networks and ISDN Systems 27(2), 165 – 173 (1994)

[16] Kesdogan, D., Agrawal, D., Pham, V., Rauterbach, D.: Fundamental Limits on the Anonymity Provided by the Mix Technique. In: Proceedings of the 2006 IEEE Symposium on Security and Privacy. pp. 86 – 99. IEEE (2006)

[17] Kesdogan, D., Pimenidis, L.: The Hitting Set Attack on Anonymity Protocols. In: Information Hiding, LNCS, vol. 3200, pp. 326 – 339. Springer (2004)
[18] Kesdogan, D., Pimenidis, L.: The Lower Bound of Attacks on Anonymity Systems – A Unicity Distance Approach. In: Quality of Protection, Advances in Information Security, vol. 23, pp. 145 – 158. Springer (2006)
[19] Köpsell, S.: Entwicklung und Betrieb eines Anonymisierungsdienstes für das WWW. Ph.D. thesis, Technische Universität Dresden (2010), german
[20] Mathewson, N., Dingledine, R.: Practical Traffic Analysis: Extending and Resisting Statistical Disclosure. In: Privacy Enhancing Technologies, LNCS, vol. 3424, pp. 17–34. Springer (2005)
[21] O'Connor, L.: Entropy Bounds for Traffic Confirmation. Cryptology ePrint Archive, Report 2008/365 (August 2008), http://eprint.iacr.org/2008/
[22] Perez-Gonzalez, F., Troncoso, C., Oya, S.: A Least Squares Approach to the Static Traffic Analysis of High-Latency Anonymous Communication Systems. Information Forensics and Security, IEEE Transactions on 9(9), 1341–1355 (Sept 2014)
[23] Pérez-González, F., Troncoso, C.: Understanding Statistical Disclosure: A Least Squares Approach. In: Privacy Enhancing Technologies, LNCS, vol. 7384, pp. 38 – 57. Springer (2012)
[24] Pfitzmann, A., Hansen, M.: Anonymity, Unobservability, Pseudonymity, and Identity Management – A Proposal for Terminology (August 2010), version v0.34
[25] Pham, D.V.: Towards Practical and Fundamental Limits of Anonymity Protection. Ph.D. thesis, University of Regensburg (November 2013)
[26] Pham, D.V., Kesdogan, D.: A Combinatorial Approach for an Anonymity Metric. In: Information Security and Privacy, LNCS, vol. 5594, pp. 26 – 43. Springer (2009)
[27] Pham, D.V., Wright, J., Kesdogan, D.: A Practical Complexity-Theoretic Analysis of Mix Systems. In: Computer Security ESORICS 2011, LNCS, vol. 6879, pp. 508 – 527. Springer (2011)
[28] Pham, V.: Analysis of the Anonymity Set of Chaumian Mixes. In: 13th Nordic Workshop on Secure IT-Systems (2008)
[29] Raymond, J.F.: Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems. In: Designing Privacy Enhancing Technologies, LNCS, vol. 2009, pp. 10–29. Springer (2001)
[30] Shannon, C.: Communication Theory of Secrecy Systems. Bell System Technical Journal 28, 656 – 715 (1949)
[31] Troncoso, C., Gierlichs, B., Preneel, B., Verbauwhede, I.: Perfect Matching Disclosure Attacks. In: Privacy Enhancing Technologies, LNCS, vol. 5134, pp. 2 – 23. Springer (2008)

A Proof of Claim

Proof (Claim 1). Let us consider the mean number of observations, such that all Alice's friends are observed at least k times exclusively, for the case that the cover-traffic is uniformly distributed. This uniform cover-traffic implies $E(T_{e,a_i}) = E(T_{e,a_j}) = E(T_e)$ for all $a_i, a_j \in {}_A\mathcal{H}$. Since the random variables $T_{k\times}$ and T_e are statistically independent, the mean number of observations until every friend is observed at least k times exclusively, equals in this case: $E(T_{k\times})E(T_e)$.

Due to the definition of $k\times$-exclusivity, observing every Alice's friend at least k times exclusively implies $k\times$-exclusivity. Therefore, we deduce the following:

$$E(T_{k\times e}) \leq E(T_{k\times})E(T_e) \ . \tag{3}$$

We now estimate $E(T_{k\times e})$ and $E(T_e)$, for arbitrary distribution of Alice's traffic and cover-traffic.

$E(T_e)$: Assume that every recipient $r \in R$, $|R| = u$ is contacted uniformly distributed by any $(b-1)$ non-Alice senders in every observation, then the probability that r is contacted by any non-Alice sender is $P_N(r) = P_N = 1-(\frac{u-1}{u})^{b-1}$. Given Alice contacts $a_j \in {}_A\mathcal{H}$ and the remaining $(b-1)$ non-Alice senders do not, then a_j is exclusive. That probability is $P_e(a_j) = (\frac{u-(m-1)}{u})^{b-1}$. The random variable T_{e,a_j} is geometrically distributed with mean:

$$E(T_{e,a_j}) = \frac{1}{P_e(a_j)} = (\frac{u-(m-1)}{u})^{1-b} \text{ , for } j = 1, \ldots, m \ . \tag{4}$$

Therefore $E(T_e) = E(T_{e,a_j})$ for all $a_j \in {}_A\mathcal{H}$, in the case of uniform cover-traffic distribution. This $E(T_e)$ serves as an upper bound for $E(T'_{e,a_j})$ of all cases, where $r' \in R'$ is non-uniformly contacted with $P'_N(r')$ and $\max_{r'\in {}_A\mathcal{H}}\{P'_N(r')\} \leq P_N$, for any recipient sets $R' \supset {}_A\mathcal{H}$.

$E(k\times)$: Let Alice contacts a friend $a \in {}_A\mathcal{H}$ (arbitrarily distributed) according to the probability mass function $P_A(a)$, where $\sum_{a\in {}_A\mathcal{H}} P_A(a) = 1$. Determining the mean number of observations $E(T_{k\times})$, until Alice contacts all her friends at least k times is equivalent to the *general coupon collector problem* (CCP) [5]. In that problem, there is a source of infinitely many coupons of the m types represented in ${}_A\mathcal{H}$, where $P_A(a)$ is the probability of drawing a coupon of type a from the source. The general CCP is to determine the mean least number of coupon collections $E(T_{k\times})$ to obtain all m coupon types.

The following equality was proved for large value of m (i.e. $m \to \infty$) in [5]:

$$E(T_{k\times}) = \frac{m}{\delta}(\ln \kappa m + \gamma) + (k-1)\frac{m}{\delta}(\ln\ln \kappa m + \ln\frac{1}{\delta}) + o(1) \ .$$

The variables in this equation have the following meaning in our context:

- $m = |{}_A\mathcal{H}|$ is the number of coupon types, where w.l.o.g. ${}_A\mathcal{H} = \{1, \ldots, m\}$.
- $\delta = \min_{x\in(0,1]} f(x) \leq 1$, where $P_A(a) = \int_{(a-1)/m}^{a/m} f(x)dx$ and $\int_0^1 f(x)dx = 1$. δ is the continuous counterpart of the discrete probability $\min_{a\in {}_A\mathcal{H}} P_A(a)$. We therefore set $f(x) = mP_A(\lceil xm \rceil)$. Therefore $\delta = m(\min_{a\in {}_A\mathcal{H}} P_A(a))$.
- $\kappa = \gamma_1 \frac{\delta^{k-1}}{(k-1)!} \leq 1$, where $0 < \gamma_1 \leq 1$ is the size of the interval, where $f(x) = \delta$.
- $o(1)$ is a negligible value.

Let $p = \min_{a\in {}_A\mathcal{H}} P_A(a)$, then $\delta = mp$. We simplify and approximate the above equation by:

$$\begin{aligned} E(T_{k\times}) &= \frac{1}{p}(\ln\frac{\gamma_1}{(k-1)!}m + \gamma) + (k-1)\frac{1}{p}\ln\ln \kappa m + o(1) \\ &\approx \frac{1}{p}(\ln m + \gamma) + (k-1)\frac{1}{p}\ln\ln m \ . \end{aligned} \tag{5}$$

The last estimate result from approximating $\frac{\gamma_1}{(k-1)!}$ and κ by its upper bound 1.

Applying the estimates (4) and (5) to inequality (3) result in (1) and completes the proof. □