On the Effectiveness of k-Anonymity Against Traffic Analysis and Surveillance

Nicholas Hopper, Eugene Y. Vasserman
University of Minnesota
200 Union St SE
Minneapolis, MN 55455 USA
hopper@cs.umn.edu, eyv@cs.umn.edu

ABSTRACT

The goal of most research on anonymity, including all currently used systems for anonymity, is to achieve anonymity through *unlinkability*: an adversary should not be able to determine the correspondence between the input and output messages of the system. An alternative anonymity goal is *unobservability*: an adversary should not be able to determine who sends and who receives messages. We study the effect of k-anonymity, a weak form of unobservability, on two types of attacks against systems that provide only unlinkability.

Categories and Subject Descriptors

C.2.0 [**Computer Networks**]: General—*Security and protection*; K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*; E.3 [**Data**]: Encryption

General Terms

Security, Theory, Measurement

Keywords

statistical disclosure, mass surveillance, k-anonymity

1. INTRODUCTION

In this paper we are concerned with two different security conditions related to anonymity:

- *Unlinkability* in an anonymity system is the property that the messages delivered by the system during some time period are "unlinkable" to the messages input to the system during that time period. Unlinkability is the typical goal of systems for anonymous communication for a number of reasons. For example, it turns out that this is exactly the property guaranteed by Chaum's mix server (if the mix is trusted and appropriate ciphertext padding is used to conceal input length) or by a mix cascade (in case at least one mix in the chain is trustworthy). Even in systems where no special trusted servers are assumed, unlinkability can typically be achieved (whether provably or not) with low communication overhead.
- *Unobservability* is a property of the principals in an anonymity scheme: roughly, a scheme is *sender unobservable* if the communications of senders and non-senders are indistinguishable, and it is *receiver unobservable* if the communications of those principals who receive messages and those who do not are indistinguishable. Receiver unobservability can be obtained, for example, by broadcast of encrypted messages, or through use of a trusted bulletin board. Sender unobservability (in a mix) can be obtained through careful use of padding and dummy messages, and is also the security goal of the DC-Net family of protocols [4, 18, 13, 10, 17]. Unobservability is not commonly provided by fielded anonymity schemes, because it involves more communication overhead, and schemes designed for unobservability may be fault-intolerant.

1.1 Attacks on unlinkable schemes

It is not hard to see that unlinkability by itself does not guarantee deterrence of all types of traffic analysis. In particular, since a system providing unlinkability need not hide the fact that Alice sends a message, or the fact that Bob receives a message, such systems can (and most often do) leak other information, such as the volume of messages sent and received by its users. This information in turn can be used effectively in several types of attacks; in this paper we specifically consider two attacks: the long-term intersection attack and the budget-constrained mass surveillance attack.

1.1.1 Intersection attacks

Long-term intersection attacks against anonymity schemes attempt to discover the pattern of communications by a single user, say Alice. As an example, consider the case where Alice sends messages through a batch mix to a single recipient, Bob. Then by intersecting the sets of users who *receive* a message from each batch where Alice *sends* a message, an eavesdropping adversary can eventually conclude that Alice's single recipient is Bob.

More general long-term intersection attacks against unlinkable anonymity schemes work under three assumptions:

WPES'06, October 30, 2006, Alexandria, Virginia, USA.

- Alice picks her recipients according to a consistent probability distribution, distinct from other users.
- The adversary can determine when a user has (with some probability) received a message
- The adversary can distinguish whether Alice might have contributed a message delivered by the system with reasonable probability.

By taking appropriate measurements of the patterns of recipients when Alice is probably sending and when she is probably not sending (not prevented solely by unlinkability), Alice's approximate set of recipients can be recovered.

An anonymity system that aims only to provide unlinkability cannot hope to prevent such an attack; even if the fact that Alice sends is concealed by some means such as running her own mix, she will occasionally go offline, allowing the attacker to label some delivered messages as not coming from Alice. Thus an important property of an anonymity system is its resistance to long-term intersection attacks, as measured by (for instance) the amount of time until an adversary can guess all of Alice's recipients.

1.1.2 Surveillance attacks

Suppose Eve wishes to eavesdrop on the communications of a large social network, by initiating a program of surveillance on the members of the network. For Eve, there is a definite *cost* to each member of the network that she places under direct surveillance - for example, the risk of getting caught infiltrating a computer, or the price of obtaining a warrant to perform the surveillance. A natural goal of Eve's would be to eavesdrop on as many members' communication as possible, given a fixed budget for surveillance. If Eve knows directly the communication patterns of each member - say, who communicates with whom - then it has been shown by Danezis and Wittneben [7] that it is possible to eavesdrop on nearly the entire network by placing only a few nodes under direct surveillance.

We would naturally expect unlinkability to increase the cost of surveillance (i.e., reduce the amount of communications eavesdropped upon for a given budget) since it conceals exactly this information. However, Danezis and Wittneben [7] have also given evidence that even given only information about the *volume* of communications produced by each member of the network (which is not necessarily concealed by a system for unlinkability), it is still possible to eavesdrop on 50% of the network members while placing only 5% of nodes under surveillance.

1.2 Effects of unobservability

In this paper, we consider the effect of unobservable communications on each of these attacks. It is easy to see that with *static* membership, that is, when all principals of the protocol are always online and participating, a receiver-unobservable anonymity scheme will prevent long-term intersection attacks completely. We extend the notion of unobservability to what we call a "periodic" model in which time is divided into fixed periods; in each period the set of principals who are online and participating is static. We show that periodic receiver-unobservability anonymity also maximally prevents long-term intersection attacks.

In the context of surveillance, Danezis and Wittneben [7] conjecture that sender-unobservability may be a better defensive measure than unlinkability. We present the results of experiments supporting this conjecture: unobservability can significantly increase the cost of achieving a given level of surveillance in a social network. Interestingly, our results suggest that unobservability is not, in itself, a perfect defense against targeted surveillance: the slope of the cost/benefit curve for surveillance against an sender - unobservable social network is still greater than the optimal defensive value of 1. We conjecture that this is a result of the "scale-free" nature of social network graphs.

Finally, we consider the effect of k-anonymity, a weaker form of unobservability, on both attacks. For the simple case of uniform background traffic, we prove that periodic receiver k-anonymity, even with a period of 1 batch, can increase the number of rounds required to find all of Alice's recipients with a given confidence level, by a multiplicative factor slightly super-linear in k, while increasing communication cost by a factor of $k/2$. We present the results of experiments showing that this gap between cost and benefit is increased when the background traffic distribution is unknown to the adversary. Finally, we present the result of experiments testing the effects of static and periodic k-anonymity on targeted surveillance of a social network. These results suggest that static, sender k-anonymity can significantly increase the cost of targeted surveillance on a social network, but that periodic k-anonymity does not provide significantly increased resistance over unlinkability.

2. RELATED WORK

Long-term intersection attacks seem to have been folklore (for instance, see [14, 3]) until the Disclosure attack was formalized by Kesdogan, Agrawal, and Penz [11]. Since carrying out the Disclosure attack requires solving an NP-hard constraint satisfaction problem, Danezis [6] proposed the Statistical Disclosure attack, which only approximately recovers the list of Alice's recipients. Mathewson and Dingledine [12] show how to extend statistical disclosure to the case of unknown background traffic, and consider the resistance to statistical disclosure provided by several factors, such as using chains of mixes, pool mixes, sender padding, and assuming incomplete network observation. They conclude that some settings of anonymity parameters can give favorable resistance to statistical disclosure. Berthold and Langos [2] note that the problem in intersection attacks is the difference in output behavior when Alice is online or offline, and propose a scheme where other network principals continue to send messages for Alice when she goes offline. It is not clear that this approach can be made practical.

Danezis and Wittneben [7] were the first to consider the issue of resistance to surveillance of social networks. They obtained the mailing list archives of an international political organization, and considered the effectiveness of various surveillance strategies. They conclude that, for their data set, unlinkable communications do not provide good protection against mass surveillance of a social network. They also suggested that unobservability might be of greater value than unlinkability in this context.

Although the majority of widely-used anonymity systems concentrate on unlinkability, several research designs for unobservable networks have been proposed. For example, the DC-Net [4] and its descendants all provide unobservable communication among a network of N principals, at a communication cost of $\Omega(N)$ per message delivered, with no trusted parties. Several other proposed protocols, including

$\mathcal{P}^5$ [15] and Xor-trees [8], provide receiver unobservability by using broadcast, which also has a worst case overhead factor of $\Omega(N)$. Proposals roughly based on mixing include Buses [1] and PipeNet [5]; these schemes can potentially reduce the communication overhead but have high latency and are vulnerable to denial-of-service attacks.

To avoid the high cost of unobservability, von Ahn, Bortz and Hopper introduced the notion of k-anonymous communication [17], analogous to the notion of k-anonymity in data privacy [16]. They observed that in many cases it could be considered sufficient if each principal is only indistinguishable from a set of $k-1$ other users, rather than all users of the network. They gave a scheme with sender and recipient k-anonymity that requires no trusted parties and has worst-case overhead $O(k^2)$.

3. PERIODIC K-ANONYMITY

In this section we review the theoretical definition of k-anonymity, and discuss why in practical terms no scheme can provide k-anonymity. We then relax this definition (which we call *static* k-anonymity) to a notion of *periodic k-anonymity* which simulates the "churn" of a realistic network. Finally, we discuss some ways in which a mix-like scheme could be extended to achieve periodic k-anonymity.

3.1 Definitions

von Ahn *et al.* [17] define an anonymous communication protocol for message space $\mathcal{M}$ as a computation among n parties $P_1, \ldots, P_N$, where each P_i starts with a series of private inputs $(msg_i, p_i) \in (\mathcal{M} \times [N]) \cup \{(\texttt{nil}, \texttt{nil})\}$, and each party terminates with a private output from $\mathcal{M}^*$. To communicate, time is split into *rounds* and after a setup round, the same transmission protocol is run at each round. Intuitively, at the end of a round each P_i should learn the set of messages addressed to him in that round ($\{msg_j : p_j = i\}$), but not the identity of the senders. For any protocol $\mathcal{P}$, we define $\mathcal{P}(P_1(x_1), P_2(x_2), \ldots, P_N(x_N))$ to be a random variable whose values are the transcripts of all communications generated when $\mathcal{P}$ is run and each P_i has private input x_i.

We now recall the definitions of sender and receiver k-anonymity [17]; Static receiver and sender unobservability can be defined as static N-anonymity.

DEFINITION 1. *A protocol $\mathcal{P}$ is* sender k-anonymous *if it induces a partition $\{V_1, \ldots, V_l\}$ of $[N]$ such that:*

1. *$|V_s| \geq k$ for all $1 \leq s \leq l$; and*
2. *For every $1 \leq s \leq l$, for all $P_i, P_j \in V_s$, for every $(msg, p) \in (\mathcal{M} \times [N]) \cup \{(\texttt{nil}, \texttt{nil})\}$, the random variables $\mathcal{P}(P_i(msg, p), *)$ and $\mathcal{P}(P_j(msg, p), *)$ are computationally indistinguishable.*

That is, each party, acting as a sender, is indistinguishable from at least $k-1$ other parties.

DEFINITION 2. *A protocol $\mathcal{P}$ is* receiver k-anonymous *if it induces a partition $\{V_1, \ldots, V_l\}$ of $[N]$ such that:*

1. *$|V_s| \geq k$ for all $1 \leq s \leq l$; and*
2. *For every $1 \leq s \leq l$, for all $P_i, P_j \in V_s$, for every $P' \in [N]$, and $msg \in \mathcal{M}$, the random variables $\mathcal{P}(P'(msg, P_i), *)$ and $\mathcal{P}(P'(msg, P_j), *)$ are computationally indistinguishable.*

That is, each message sent to an honest party has at least k indistinguishable recipients.

Notice that a static k-anonymous protocol will protect a sender or receiver absolutely as long as the set of recipients remains the same - he will always be indistinguishable from at least $k-1$ other participants. However, real-life network protocols experience *churn*: new members will join the network, old members will leave, and existing members may not always be able to participate. To deal with this situation, we introduce the notion of *periodic k-anonymity*. A periodic anonymity protocol has a pool of participants $\{Q_1, \ldots, Q_n\}$ out of which N parties – $P_1, \ldots, P_N$ – participate in each round. A periodic protocol should have the property that if the set S of participants in two rounds is the same, then the partition $\mathcal{V}(S) = \{V_1, \ldots, V_l\}$ remains the same; if the set of participants is different, then we assume that the partitions $\mathcal{V}_1, \mathcal{V}_2$ are independent.[1]

In such a setting, which is essentially pessimistic for our results, a critical security parameter (controlled by the environment the protocol is used in, but not necessarily by the deployer) is the *churn rate*. We say that an environment has churn rate ρ if the set of participants only changes once every ρ rounds. Low values of ρ mean that users will change groups often and thus not "blend in" as well, while higher values of ρ require a more stable network.

3.2 k-anonymity with a Mix

Although the protocol of [17] has "low overhead" in theoretical terms, it is not very practical for real deployment. Much of this is due to the fact that the protocol is intended to avoid the use of trusted servers. Here we give short sketches of how a mix server could be modified to provide periodic k-anonymity for senders or receivers while requiring fairly minimal effort on the part of the receivers. Note that it is not our intent to describe a complete system, as here we are more interested in studying the effect of combining k-anonymity with unlinkable systems than the exact details of any particular protocol; From the point of view of the disclosure and surveillance attacks, the anonymity scheme is essentially a black box.

Receiver k-anonymity. Suppose a mix has a list of possible message recipients. This list could be obtained in a number of ways - for example, requiring a sender to (anonymously) register any message recipient at least one round before he will send a message; or just by adding an address to the list when the mix sees a message with that destination address. It is easy for the mix to partition this list of recipients into groups so that each recipient has at least $k-1$ others in his group. If each recipient has a public key, then when processing a batch, the mix can encrypt dummy messages to send to each member of a group who does not have a message in the batch, so that all k members of each group receive the same number of messages out of each batch.

Sender k-anonymity. k-anonymity for senders could be

[1]We note that it would be *desirable* to engineer protocols so that small changes in the set of participants leave many sets of a partition unchanged, since this would further limit the effectiveness of intersection attacks. If real-world protocols are engineered in this way, our results will thus *understate* the effectiveness of k-anonymity against intersection attacks.

achieved in several ways. For example, the mix could require senders to enroll (with a public key), partition senders into groups of at least k members, and then periodically send "tokens" to all members of a group, requiring them all to reply with either an encrypted dummy message or an outgoing message before processing. Other alternatives are possible if special software is employed, for example composing the mix with the k-AMT protocol of von Ahn *et al.* or the unobservable protocol of Golle and Juels [10].

4. STATISTICAL DISCLOSURE AGAINST K-ANONYMITY

Mathewson and Dingledine [12] show that the statistical disclosure attack is in fact a very general attack which can defeat many different types of mix networks as well as many network countermeasures which protect against short-term pattern analysis. In this attack, an adversary, without prior knowledge of network conditions, attempts to probabilistically identify Alice's recipients based only on Alice's send pattern and the receive pattern of other nodes in the network (whose received messages are composed of Alice's messages and background traffic from nodes other than Alice). This attack is effective against many different types of mix networks, but especially against those where variability in the timing of message delivery is fairly low. In fact, the attack is so strong that it's possible to treat the entire mix network as a black box while only concerning ourselves with the pattern of how messages enter and leave the network.

The attack fails in only a few cases, such as when Alice's behavior is unpredictable, when the attacker can not observe that Alice is sending messages, or when the attacker can not observe how the network behaves at a time when Alice is not sending messages.

4.1 Basic attack

The basic statistical disclosure attack [6] models Alice as having m possible recipients. Every time Alice sends a message, she picks a recipient from that set with probability $\frac{1}{m}$, and sends a message to the batch mix, addressed to the selected recipient. The mix then receives $b-1$ other messages addressed to recipients chosen uniformly from the N participants in the network. The attack model's Alice's recipients as a probability distribution $\overrightarrow{v}$, with the value $\frac{1}{m}$ in the locations corresponding to Alice's recipients, and 0 elsewhere. We let the vector $\overrightarrow{u}$ model the background traffic of the network in a similar manner, i.e. in the basic model we assume that $\overrightarrow{u}$ is an N-element vector whose elements are all $\frac{1}{N}$.

The adversary constructs a vector $\overrightarrow{o}$ which models the behavior of the network during any given round, such that every recipient that received a message during a round Alice sent a message has a value of 1, while others have the value of 0.

$$\overline{O} = \frac{1}{t}\sum_{i=i}^{t}\overrightarrow{o_i} \approx \frac{\overrightarrow{v} + (b-1)\overrightarrow{u}}{b} \ .$$

We can then estimate $\overrightarrow{v}$ from the arithmetic mean of many round vectors $\overrightarrow{o}$:

$$\overrightarrow{v} \approx b\frac{\sum_{i=1}^{t}\overrightarrow{o_i}}{t} - (b-i)\overrightarrow{u} \ .$$

Danezis [6] uses basic signal processing techniques to show that in order to determine Alice's recipients with 95% confidence, it is sufficient to observe t rounds in which Alice sends a message, where

$$t > \left[2m\left(\sqrt{\frac{N-1}{N}(b-1)} + \sqrt{\frac{N-1}{N^2}(b-1) + \frac{m-1}{m}}\right)\right]^2 .$$

4.2 Extended Attack

Mathewson and Dingledine strengthen the Danezis intersection attack to more closely match real-world mix networks. Under the updated scheme, Alice is allowed to send more than one message in any given round, and she can select recipients (from a set of possible recipients) in a non-uniform way. Also, the adversary is no longer required to have complete knowledge of the background traffic distribution.

To carry out this attack, the adversary must first gather a good sample of the network background during rounds when Alice is not sending messages (the background distribution estimate is similar to that in the original attack). The attacker then observes rounds in which Alice participates (sends messages), and computes the arithmetic mean of the probability vectors for each round as before. From this information, the attacker computes an estimate of Alice's behavior:

$$\overrightarrow{v} \approx \frac{1}{\overline{m}}\left[b\cdot\overline{O} - (b-\overline{m})\overline{U}\right]$$

where $\overline{U}$ is the estimate of the unknown background traffic.

Mathewson and Dingledine show the effectiveness of this attack against the standard mix network, even with the unknown background adversary and the weighted Alice. The results for pool mix networks showed that the intersection attack is very effective in low delay probability networks for any message volume from Alice (being more effective at intermediate message volumes, and less effective at very low or very high message volumes), but fails at higher delay probabilities and low/high message volume from Alice, although message volumes of 0.5-0.6 still yield a successful attack in all cases.

Mathewson and Dingledine further extend the attack to deal with pool mixes, sender dummy traffic, time-variant background traffic, and partial observability. We do not consider these measures here as we are focused mainly on the effectiveness of k-anonymity.

4.3 Theoretical results

In this section, we develop two theoretical results about the effects of periodic k-anonymity on the statistical disclosure attack. First, we prove that even 1-periodic receiver N-anonymity prevents statistical disclosure, up to revealing information available to the attacker. Thus if we are willing to tolerate high communication overhead, statistical disclosure attacks can be essentially perfectly resisted. Second, we derive a bound on the expected number of rounds necessary to execute the statistical disclosure attack against a ρ-periodic k-anonymous mix with uniform background traffic, and show that this figure is super-linear in k, and at least affine in ρ. Thus, higher values of k and a higher "churn period" give better resistance to statistical disclosure.

4.3.1 Periodic Unobservability prevents long-term intersection

Suppose we have a periodic unobservable anonymity protocol in which all participants take part whenever they are online. We would like to prove that no intersection attack is possible on such a scheme - intuitively, the adversary can never tell which participants are sending or receiving messages. Of course, no matter what protocol the participants run, the adversary will still be able to see when the participants are online and when they are offline. It is certainly possible to construct sequences of participant joins and leaves so that this information leaks the communication patterns of the participants, but no anonymity scheme could prevent such leakage, since any reasonable model of a global adversary would allow him to see which participants are online and offline (for example, by noting the presence or absence of other traffic to the participants, or by "pinging" them). In practical situations, however, we would expect that most users do not show any noticeable correlation in online status with their correspondents that they do not also show with, e.g., other users on the same approximate sleep schedule (in fact, they may show stronger correlation to these users than their correspondents). The following theorem uses proof techniques developed in the cryptographic literature to show that given the list of participants in each period, an adversary learns no further information about the communication patterns of the participants from a periodic sender and receiver unobservable anonymity scheme.

Let $\overrightarrow{\mathcal{Q}} = \langle \mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_t \rangle$ be a sequence of participant sets for a periodic unobservable networking protocol $\mathcal{P}$; let $\overrightarrow{R} = \langle R_1, R_2, \ldots, R_t \rangle$ be a sequence of recipients such that $R_i \in \mathcal{Q}_i \cup \texttt{nil}$, and let $\mathcal{P}_{alice}(\overrightarrow{R}, \overrightarrow{\mathcal{Q}})$ denote the result of running $\mathcal{P}$ with Alice sending encrypted messages to R_i at round i, with participants $\mathcal{Q}_i$. The goal of a long-term intersection attack is to discover some information about $\bigcup_i \{R_i\}$ given $\mathcal{P}_{alice}(\overrightarrow{R}, \overrightarrow{\mathcal{Q}})$. We show that if Alice follows the protocol, then *even when an attacker knows the recipient distribution of all other senders*, whatever he can learn given $\mathcal{P}_{alice}(\overrightarrow{R}, \overrightarrow{\mathcal{Q}})$, he can also learn given only $\overrightarrow{\mathcal{Q}}$. Since an adversary can always learn what principals are online at a given time (for example, by use of "ping") this implies that periodic unobservability gives essentially the best possible resistance to long-term intersection attacks.

THEOREM 1. *Under the conditions above, for every efficient "disclosure" attacker A, there is a simulated attacker S such that $S(\overrightarrow{\mathcal{Q}})$ and $A(\mathcal{P}_{alice}(\overrightarrow{R}, \overrightarrow{\mathcal{Q}}))$, are computationally indistinguishable.*

PROOF. The proof is straightforward. Imagine an alternative universe in which, for every round that Alice is online, she sends a message to the first participant in $\mathcal{Q}_i$; let this recipient list be denoted $\overrightarrow{R'}$. Certainly, seeing the result of this protocol execution will not reveal anything about Alice's original recipient list $\overrightarrow{R}$, since $\overrightarrow{R'}$ is uncorrelated to $\overrightarrow{R}$, given $\overrightarrow{\mathcal{Q}}$. We will prove, however, that if the anonymity scheme is periodically sender unobservable, then the transcripts with these two sets are computationally indistinguishable: anything you can efficiently learn from the first transcript (given $\overrightarrow{\mathcal{Q}}$) you can efficiently learn from the second as well.

Now imagine changing the list $\overrightarrow{R}$ into $\overrightarrow{R'}$ one element at a time: first change R_1 to R'_1, then change R_2 to R'_2, and so on. Call these vectors $\overrightarrow{R_1}, \overrightarrow{R_2}, \ldots, \overrightarrow{R_t}$. Now the only difference between $\mathcal{P}_{alice}(\overrightarrow{R}, \overrightarrow{\mathcal{Q}})$ and $\mathcal{P}_{alice}(\overrightarrow{R_1}, \overrightarrow{\mathcal{Q}})$ is who Alice sends to in the first round; but by receiver unobservability, we know that the transcript of this first round should be indistinguishable whether Alice sends to R_1 or R'_1. Similarly, the only difference between $\mathcal{P}_{alice}(\overrightarrow{R_i}, \overrightarrow{\mathcal{Q}})$ and $\mathcal{P}_{alice}(\overrightarrow{R_{i+1}}, \overrightarrow{\mathcal{Q}})$ is the recipient in round $i+1$, and since the transcript in round $i+1$ with R_i or R'_i as recipients should be indistinguishable by receiver unobservability, the entire transcripts should also be indistinguishable. Extending this argument t times, we should conclude that $\mathcal{P}_{alice}(\overrightarrow{R}, \overrightarrow{\mathcal{Q}})$ and $\mathcal{P}_{alice}(\overrightarrow{R'}, \overrightarrow{\mathcal{Q}})$ are indistinguishable; this can be formalized as the standard "hybrid argument" from the cryptographic literature (see, for example [9]).

Formally, we let $\overrightarrow{R'} = \langle R'_1, R'_2, \ldots R'_t \rangle$ be any vector of recipients such that $R_i \in \mathcal{Q}_i \setminus \{alice\}$ when $alice \in \mathcal{Q}_i$ and $R_i = \texttt{nil}$ otherwise. Then by the definition of unobservability and a standard hybrid argument, $\mathcal{P}_{alice}(\overrightarrow{R}, \overrightarrow{\mathcal{Q}})$ and $\mathcal{P}_{alice}(\overrightarrow{R'}, \overrightarrow{\mathcal{Q}})$ are indistinguishable. But since $\overrightarrow{R'}$ can be computed without knowing Alice's private input $\overrightarrow{R}$, Alice follows the protocol, and the simulator knows the "background" distribution of other senders, S can simply simulate a complete run of the protocol using the participant sets $\overrightarrow{\mathcal{Q}}$, computing messages that Alice would output with $\overrightarrow{R'}$ as recipients and dummy messages as content. S can then run A on this simulated transcript and output the result, and since this input will be computationally indistinguishable from $\mathcal{P}_{alice}(\overrightarrow{R}, \overrightarrow{\mathcal{Q}})$, the output of S will be computationally indistinguishable from $A(\mathcal{P}_{alice}(\overrightarrow{R}, \overrightarrow{\mathcal{Q}}))$. □

4.3.2 Bounds on statistical disclosure with periodic k-anonymity

As in the original statistical disclosure attack, we can derive, for a uniform background, the expected number of rounds to distinguish the support of Alice's recipient distribution from a noise coordinate, when using ρ-periodic k-anonymity. We assume that in each round, Alice sends a message to one of her m recipients chosen uniformly at random, and $(b-1)$ other messages are delivered to recipients chosen uniformly among the N participants. Furthermore, each message delivered to a participant also causes $k-1$ dummy deliveries to the participants in the same partition, so that the expected number of messages delivered to each non-recipient as a result of these inputs is $(b-1)\frac{k}{N}$.
We let l be a confidence parameter, and wish to solve for the number of rounds t such that

$$\mu_{alice} - \mu_{noise} > l(\sigma_{alice} + \sigma_{noise}) \ .$$

Using the approximation $1-(1-\frac{1}{N})^k \approx \frac{k}{N} \approx \frac{k-1}{N} = \varepsilon$, which is valid for $k \ll N$, we find that this requires

$$t > l^2 m^2 \left[\begin{array}{c} \sqrt{\varepsilon\left(b(1-\varepsilon) + (\rho-1)(\frac{1}{m}-\varepsilon)\right)} \\ + \sqrt{\begin{array}{c} f(1-f\rho) + (b-1)\varepsilon(1-\varepsilon) \\ + (\rho-1)(\frac{1}{m^2} + \frac{(1+4m)(m-1)\varepsilon}{m^2}) \end{array}} \end{array} \right]^2 ,$$

where $f = \frac{N+(m-1)(k-1)}{mN}$. It can be verified that this bound is better than linear in $\frac{k}{N}$, for small values of $\frac{k}{N}$, and essentially linear in the batch size b and the churn rate ρ.

4.4 Simulation results

We modify the mixnet simulator from [12] to optionally apply the k-anonymity transform to any currently-running

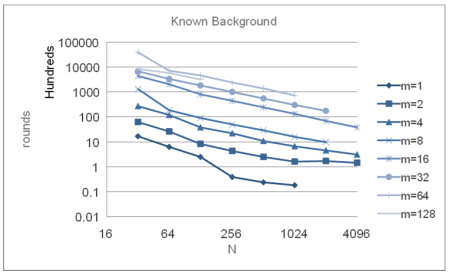


Figure 1: Known background network with k-anonymous group size of 16.

experiment. We simulate network churn by using static k-anonymity for a number of rounds (ρ), then "churning" the entire mix by breaking up all k-anonymous groups and dynamically reforming them later, as needed. Groups are formed when a recipient not yet assigned to a group gets a message. That recipient is then assigned to a group with up to k nodes, uniformly selected from all nodes that are not members of a group. This is a gross simplification of the way churn would work in a real network, in which individual nodes in a given k-anonymous group have a given probability of leaving that group in any given round, either voluntarily or due to failure.

The k-anonymity module works by modifying the traffic vector $\overrightarrow{t}$ using an offset vector $\overrightarrow{o}$. The traffic vector represents the entire network message history (who received messages, and how many), and the offset vector represents the network activity during any given round. For every message $M \in \overrightarrow{o}$ we copy the message into $\overrightarrow{t}$, destined to the original recipient R, and also add up to k copies destined to every node in R's k-anonymous group, thus increasing the number of messages in the network by at most a factor of $\frac{k+1}{2}$ [2].

In order to confirm that our modified simulator preserves the behavior of the original, we run the known background simulation and apply the k-anonymity transform ($k = 16$) with variable number of recipients. Figure 1 shows that the behavior of the modified simulator is consistent with the behavior of the original simulator (compare to Figure 1 in [12]).

Figure 2 shows that we achieve an anonymity improvement that is larger than can be accounted for simply by the increased number of messages in the network (the number of rounds required for a successful intersection attack grows faster than k, with best anonymity improvements visible at higher k values): introducing k-anonymity into a standard mixnet will increase the difficulty of the statistical disclosure attack (on average) by a factor of $k^{1.15}$ at a churn rate of once every 60 rounds.

We additionally examine the behavior of k-anonymity together with complex sender behavior and unknown background traffic. For these experiments, we model the background traffic and sender traffic, or just the background traffic, according to a small-world network model [20], which is meant to represent real-world networks such as social networks and the Internet, where the majority of individuals can communicate with each other just by maintaining connections to a small number of other individuals. When the sender or background traffic are not assumed to be "complex," Alice or the background select recipients uniformly from the appropriate set. Figure 3 shows that the privacy-enhancing effect of k-anonymity is retained even in complex mixnet models. Once again, the privacy benefit is slightly super-linear in k with a traffic volume increase of at most a factor of $\frac{k+1}{2}$. Note that in (a), the line for $m = 16$ flattens at $k \geq 8$ because attacks fail to succeed before the cutoff of 5000000 rounds.

[2]The increase in the number of messages is slightly smaller, in situations where k does not divide N without remainder, in which case at least one group of size $< k$ will be formed.

5. RESISTING MASS SURVEILLANCE

Consider the problem of concealing the fact that Alice and Bob communicate with each other. Cryptography can hide the contents of their communications, and anonymity can (perhaps) hide the fact that they communicate (for some time), when only their communications are under observation. However, in most cases Alice and Bob will not only communicate with each other, but will be part of a larger "social network" [19]. Thus we would expect that Alice and Bob's correspondence may well be leaked in their dealings with others, so that even if Alice and Bob are not compromised, compromising other nodes in the network may reveal their existence and correspondence as well.

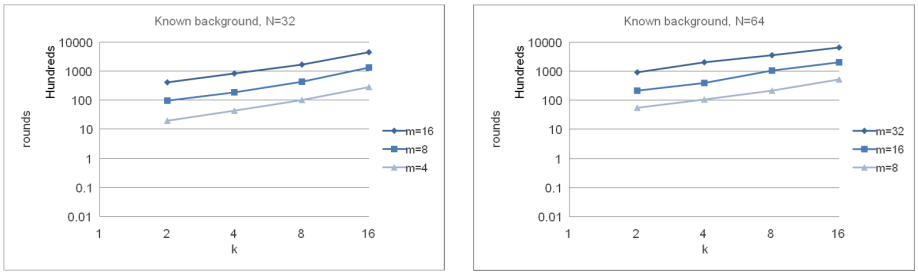
5.1 Setting

Danezis and Wittneben [7] introduce the following natural problem in this setting: given that such externalities exist, and that there is a definite cost in compromising each member of a social network – for example, the risk of being caught, or the cost of obtaining a warrant, plus perhaps the cost of finding and exploiting a security weakness in a given machine – how can an adversary maximize the number of correspondences learned, or the number of individuals under "indirect" surveillance for a given surveillance budget? Or, from the point of view of security, how can we maximize the resistance of a social network to such surveillance efforts?

We follow [7] in modeling a social network as a set of *individuals*, who are connected to *spaces*. The reader may think of a space as a club or committee; communication between its members is assumed to form a clique. The link between an individual and a space has an associated weight, which indicates the volume of traffic the individual sends to the space. The problem then becomes, given partial information about the members of the network, to maximize the number of individuals or spaces under surveillance for a given number of compromised individuals.

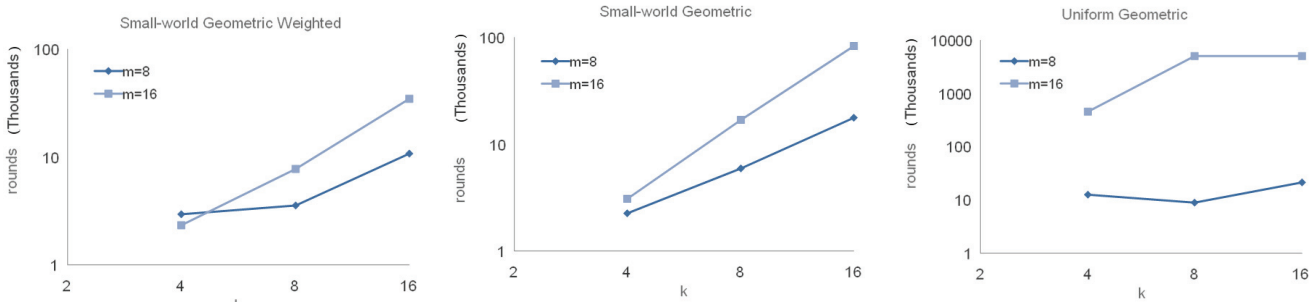
Danezis and Wittneben study the efficacy of many surveillance strategies given full information about a social network and given only the total volume of communications by each node in the network – exactly the information leaked by an unlinkable anonymity scheme such as a mix network. They use as a basis for their study a social network constructed from the mailing list archive of an international political organization; their data set spans roughly three years, and consists of 2338 people in 373 spaces.

The authors of [7] find that, given full information on the communication patterns of a network, a very small number of compromises can lead to surveillance of essentially the entire network. An adversary who first compromises the invididual connected to the most spaces, and subsequently



(a) (b)

Figure 2: Statistical disclosure results against a known background distribution for varying levels of k with (a) $N = 32$ and (b) $N = 64$.



(a) (b) (c)

Figure 3: Experiments using attack adapted to unknown background distribution. (a) Small-world geometric network network model used for both background and sender distributions; (b) Small-world geometric model background, uniform sender distribution; (c) Uniform background and sender distributions.

compromises individual that maximize the new number of spaces under surveillance, can uncover 100% of the spaces while compromising only 8% of the individuals in the network, and can uncover 50% of spaces by compromising less than 2% of the individuals.

To assess the effectiveness of unlinkable anonymity against mass surveillance, the authors simulated an adversary who first compromises the individual with the highest total volume of traffic, and subsequent compromises the nodes with the highest remaining volume of traffic. This strategy, shown to be more effective than an adaptive strategy, uncovers 50% of spaces with only 5% of nodes compromised. Thus, while unlinkability moderately improves resistance to mass surveillance, it does not do so dramatically. [7] conjecture that unobservability is needed to provide better resistance against mass surveillance attacks.

5.2 Our results

To test the resistance of unobservability, k-anonymity, and periodic k-anonymity against mass surveillance, we obtained three public data sets of a nature similar to those used by [7]:

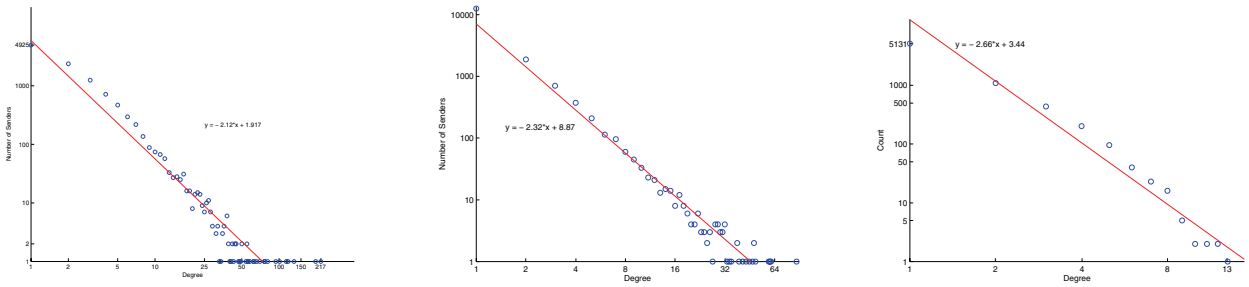
- `ietf`: the public archive of the mailing lists of the Internet Engineering Task Force, since January 1, 2000. This network consisted of 10978 individuals participating in 305 spaces.
- `w3c`: the complete public archive of the mailing lists of the World Wide Web Consortium. This data set had 16644 individuals participating in 269 spaces, with the earliest posts occurring in June 1994, up until June 2006.
- `hwg`: the complete public archive of the mailing lists of the HTML Writers' Guild. This data set had 4418 individuals participating in 19 spaces, with the earliest postings in December 1997 and the latest posting in February 2005.

Figure 4 shows the degree distributions of the individuals of each data set. As expected, all three data sets match a power law distribution, as in the data sets of [7].

We performed three sets of experiments. The first experiment measured the surveillance resistance of each data set with no anonymity scheme, with an unlinkable scheme, and with unobservability. The second experiment measured the effect of static k-anonymity on surveillance resistance for various levels of k. Finally, we measured the effect of periodic churn with various levels of k-anonymity.

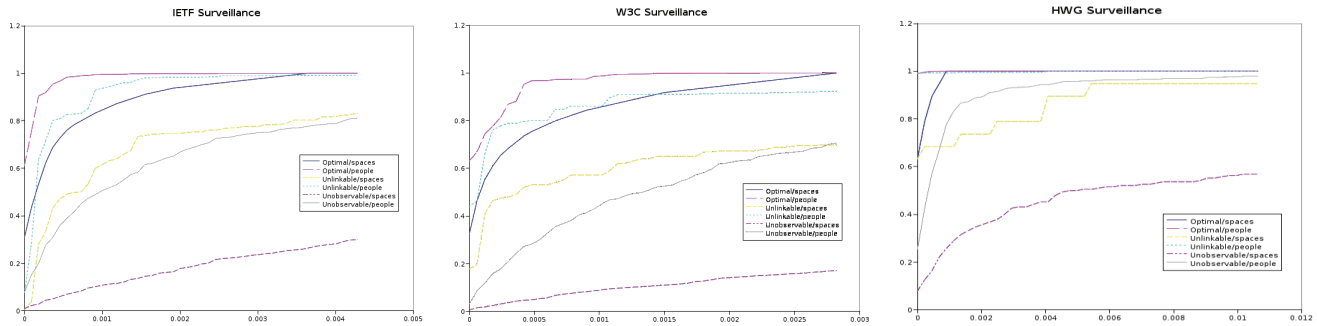
5.2.1 *Full information, unlinkability, and unobservability*

For each data set, we simulated the attack based on full information and the attack based on volume information only, as described in [7]. Our results confirm the experiments reported there; in fact, surveillance is even more effective on our data sets. For instance, in the `ietf` data set, with full information, all spaces are uncovered with only 41 individuals, or 0.4%, under direct surveillance and 50% of spaces are uncovered with only 3 individuals under direct surveillance. In the same data set, with only volume information, 90% of spaces are uncovered with only 147 individuals under direct surveillance, or about 1.4%, while 50% of spaces are uncovered with only 9 individuals under direct surveillance.



(ietf) (w3c) (hwg)

Figure 4: Degree distribution plots for the three data sets used in our experiments. All data sets follow a power-law distribution, as expected in social networks.



(ietf) (w3c) (hwg)

Figure 5: Results of surveillance (people and spaces uncovered) for optimal surveillance, unlinkable surveillance, and unobservable surveillance, for each data set.

An anonymity scheme with static unobservability would give an adversary no information, aside from a list of probable members, about the connections between them; thus the only surveillance strategy available would be to randomly select individuals for direct surveillance. We simulated 10 random trials of this attack against each data set. Our results suggest that unobservability may still be insufficient as a defense against mass surveillance. For example, in the `ietf` data set, on average 50% of spaces are uncovered with 127 individuals, or 1.2%, under direct surveillance and 90% of spaces are uncovered with 1434 individuals, or less than 15%, under direct surveillance. While this is a significant improvement in resistance over unlinkability (for our dataset), it is still far from the ideal situation in which we would hope to require 50% of individuals under direct surveillance to uncover 50% of spaces. Our results for all data sets are summarized in Figure 5.

5.2.2 *Effects of k-anonymity and churn*

To assess whether static, unlinkable k-anonymity could offer resistance similar to unobservability, we simulated 10 runs (with randomly chosen partitions of senders) of the unlinkability attacker against a k-anonymized data set for $k \in \{2, 8, 32\}$. In these trials, every time some Alice in our data set sent a message to a mailing list, we incremented the "apparent volume" of each member of her k-individual partition; the attacker then selected individuals to place under surveillance by decreasing "apparent volume."

Our results show that k-anonymity can offer some benefit over unlinkability at a much lower communication overhead than static unobservability. For example, with $k = 2$, in the `ietf` data set, 50% of spaces are uncovered with 14 individuals under direct surveillance (opposed to 9), and for $k = 32$ the number of individuals required to uncover 50% of spaces on average was 77. For $k = 32$, the average number of individuals required to be under direct surveillance to uncover 90% of spaces was 845. Full results of this experiment appear in Figure 6.

To simulate the effects of churn on the resistance provided by k-anonymity, we divided time into periods of one day, one week, one month, and one year. An individual was considered to be a participant for a given time period if he/she contributed at least one posting during that time period. For each time period, the set of participants were grouped into randomized partitions of size at least k, and each time an individual in a partition sent a posting, all k members of the partition had their "apparent traffic volume" incremented. We then simulated the attacker who selects individuals in decreasing order of apparent volume.

Our results show that even a modest rate of churn (once a year) will significantly aid the attacker. For example, with the `ietf` data set, when $k = 32$, the average number of individuals under surveillance required to uncover 50% of spaces with churn of one week, one month, and one year were 5, 6, and 46, respectively. To uncover 90% of spaces, the required number of individuals were 173, 189, and 379, respectively. Results for the experiment with $k = 8$ on all three data sets are shown in Figure 7.

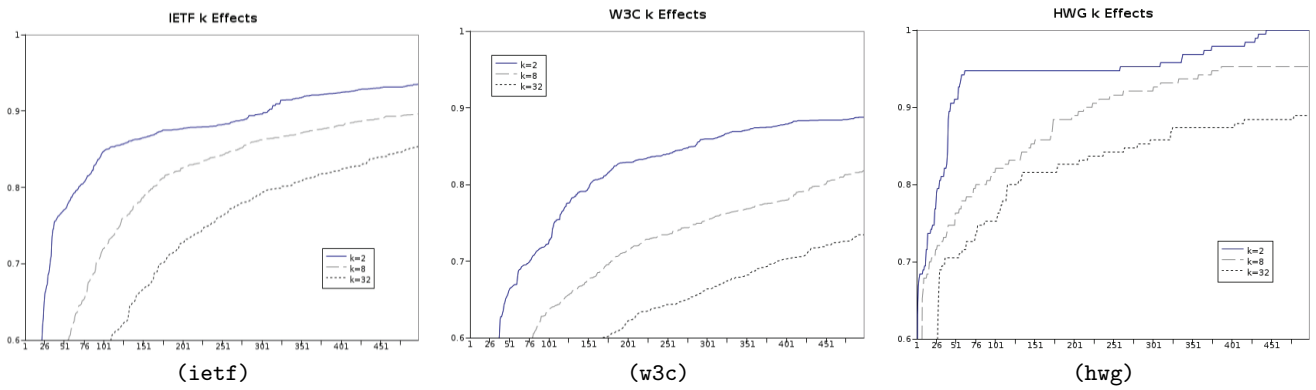


Figure 6: Results of surveillance, (fraction of spaces uncovered) with static k anonymity, for $k \in \{2, 8, 32\}$

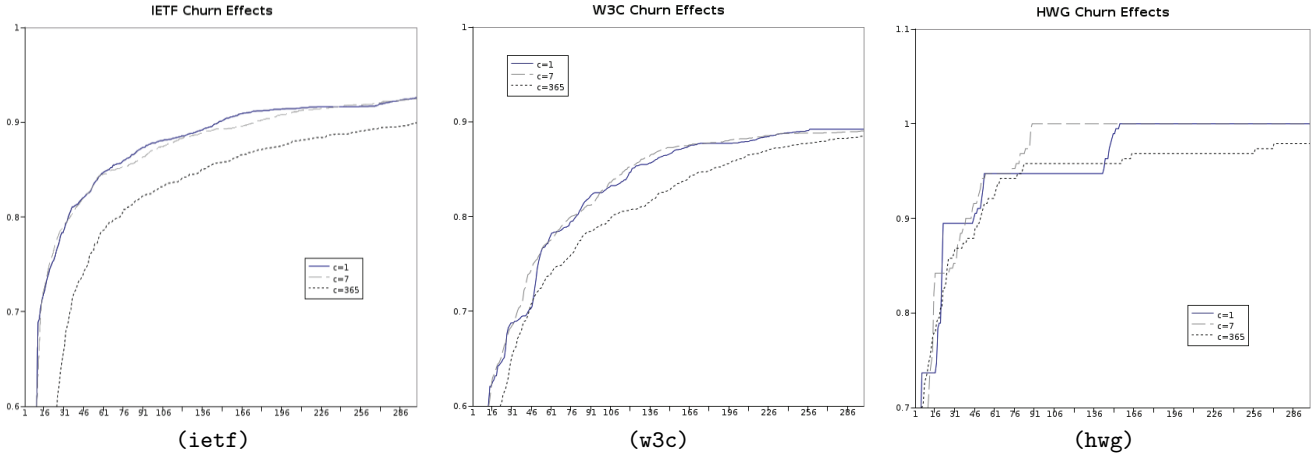


Figure 7: Results of surveillance, (fraction of spaces uncovered) with periodic 8-anonymity, for churn of one day, one week, and one year.

6. CONCLUSIONS

Regarding resistance to statistical disclosure, our results suggest that receiver k-anonymity can strengthen a batch mix against long-term intersection attacks, for low values of k. This strengthening is improved when the list of willing recipients of a mix has a higher churn interval. It is not a "time-space tradeoff" in the sense that increasing traffic by a factor of $k/2$ improves resistance to the disclosure attack by a minimum factor of $k^{1.15}$.

Regarding mass surveillance, we conclude that unobservability and periodic sender k-anonymity in environments with low churn can significantly improve resistance to mass surveillance (the value of such systems is less "questionable" than in the case of unlinkable anonymity schemes), but still does not approach the optimal cost function.

In terms of resistance to intersection-type attacks, our work raises several interesting questions for future research. For example, how would receiver k-anonymity interact with the resistance provided by various techniques explored in [12] such as pool mixes and sender padding? What is the effect of sender k-anonymity on statistical disclosure attacks? It would also be of interest to find good estimators of the difficulty of statistical disclosure with k-anonymity when the background traffic is unknown.

In terms of resisting mass surveillance, an interesting question is how the parameters of the social network influence the value of various anonymity mechanisms. We also expect that it will be challenging to find new techniques that protect social networks from this type of leakage.

7. ACKNOWLEDGEMENTS

The authors thank Luis von Ahn, Roger Dingledine, Yongdae Kim, and the anonymous WPES referees for their helpful comments and discussions about the paper. This research was supported by the US National Science Foundation under grant CNS-0546162 and by the University of Minnesota's Grant-in-Aid of Research, Artistry and Scholarship Program.

8. REFERENCES

[1] A. Beimel and S. Dolev. Buses for anonymous message delivery. *Journal of Cryptology*, 16(1):25–39, 2003.

[2] O. Berthold and H. Langos. Dummy traffic against long term intersection attacks. In R. Dingledine and P. Syverson, editors, *Proceedings of Privacy Enhancing Technologies workshop (PET 2002)*. Springer-Verlag, LNCS 2482, April 2002.

[3] O. Berthold, A. Pfitzmann, and R. Standtke. The disadvantages of free MIX routes and how to overcome

them. In H. Federrath, editor, *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, pages 30–45. Springer-Verlag, LNCS 2009, July 2000.

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

[5] W. Dai. Pipenet 1.1. Usenet post, August 1996.

[6] G. Danezis. Statistical disclosure attacks: Traffic confirmation in open environments. In Gritzalis, Vimercati, Samarati, and Katsikas, editors, *Proceedings of Security and Privacy in the Age of Uncertainty, (SEC2003)*, pages 421–426, Athens, May 2003. IFIP TC11, Kluwer.

[7] G. Danezis and B. Wittneben. The economics of mass surveillance and the questionable value of anonymous communications. In *Proceedings of the 5th Workshop on The Economics of Information Security (WEIS 2006)*, June 2006.

[8] S. Dolev and R. Ostrobsky. Xor-trees for efficient anonymous multicast and reception. *ACM Trans. Inf. Syst. Secur.*, 3(2):63–84, 2000.

[9] O. Goldreich. *Foundations of Cryptography: Basic Tools*. Cambridge University Press, New York, NY, USA, 2000.

[10] P. Golle and A. Juels. Dining cryptographers revisited. In *Proceedings of Eurocrypt 2004*, May 2004.

[11] D. Kesdogan, D. Agrawal, and S. Penz. Limits of anonymity in open environments. In F. Petitcolas, editor, *Proceedings of Information Hiding Workshop (IH 2002)*. Springer-Verlag, LNCS 2578, October 2002.

[12] N. Mathewson and R. Dingledine. Practical traffic analysis: Extending and resisting statistical disclosure. In *Proceedings of Privacy Enhancing Technologies (PET 2004)*, volume 3424 of *LNCS*, May 2004.

[13] A. Pfitzmann and M. Waidner. Networks without user observability – design options. In *Proceedings of EUROCRYPT 1985*. Springer-Verlag, LNCS 219, 1985.

[14] J.-F. Raymond. Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems. In H. Federrath, editor, *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, pages 10–29. Springer-Verlag, LNCS 2009, July 2000.

[15] R. Sherwood, B. Bhattacharjee, and A. Srinivasan. P5: A protocol for scalable anonymous communication. In *Proceedings of the 2002 IEEE Symposium on Security and Privacy*, May 2002.

[16] L. Sweeney. k-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570, 2002.

[17] L. von Ahn, A. Bortz, and N. J. Hopper. k-anonymous message transmission. In V. Atluri and P. Liu, editors, *Proceedings of the 10th ACM Conference on Computer and Communications Security (CCS 2003)*, pages 122–130. ACM Press, October 2003.

[18] M. Waidner and B. Pfitzmann. The dining cryptographers in the disco: unconditional sender and recipient untraceability with computationally secure servicability. In *Proceedings of EUROCRYPT 1989*. Springer-Verlag, LNCS 434, 1990.

[19] S. Wasserman, K. Faust, D. Iacobucci, and M. Granovetter. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1995.

[20] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, June 1998.