Probabilistic Analysis of Onion Routing in a Black-box Model

JOAN FEIGENBAUM, Yale University
AARON JOHNSON, U.S. Naval Research Laboratory
PAUL SYVERSON, U.S. Naval Research Laboratory

We perform a probabilistic analysis of onion routing. The analysis is presented in a black-box model of anonymous communication in the Universally Composable framework that abstracts the essential properties of onion routing in the presence of an active adversary who controls a portion of the network and knows all *a priori* distributions on user choices of destination. Our results quantify how much the adversary can gain in identifying users by exploiting knowledge of their probabilistic behavior. In particular, we show that, in the limit as the network gets large, a user u's anonymity is worst either when the other users always choose the destination u is least likely to visit or when the other users always choose the destination u chooses. This worst-case anonymity with an adversary that controls a fraction b of the routers is shown to be comparable to the best-case anonymity against an adversary that controls a fraction $\sqrt{b}$.

Categories and Subject Descriptors: C.2.0 [**Computer-Communication Networks**]: General—*security and protection*; C.2.4 [**Computer-Communication Networks**]: Distributed Systems—*Distributed applications*; K.4.1 [**Computers and Society**]: Public Policy Issues—*privacy*; G.3 [**Probability and Statistics**]: *probabilistic algorithms*

General Terms: Security, Theory

Additional Key Words and Phrases: anonymous communication, onion routing, Tor

1. INTRODUCTION

Every day, half a million people use the onion-routing network Tor [Dingledine et al. 2004] to anonymize their Internet communication. However, the effectiveness of this service, and of onion routing in general, is not well understood. The approach we take to this problem is to model onion routing formally all the way from the protocol details to the behavior of the users. We then analyze the resulting system and quantify the anonymity it provides. Key features of our model include i) a black-box abstraction in the Universally Composable (UC) framework [Canetti 2000] that hides the underlying operation of the protocol and ii) probabilistic user behavior and protocol operation.

Systems for communication anonymity generally have at most one of two desirable properties: provable security and practicality. Systems that one can prove secure have used assumptions that make them impractical for most communication applications. Practical systems are ultimately the ones we must care about, because they are the ones that will actually be used. However, their security properties have not been rig-

Joan Feigenbaum (email: Joan.Feigenbaum@yale.edu) was supported in part by NSF grants 0331548 and 0534052, ARO grant W911NF-06-1-0316, and US-Israeli BSF grant 2002065. Aaron Johnson (email: aaron.m.johnson@nrl.navy.mil) did the majority of this work while at Yale University and was supported by NSF grant 0428422 and ARO grant W911NF-05-1-0417. Some work was also done while at The University of Texas at Austin. Paul Syverson (email: syverson@itd.nrl.navy.mil) was supported by ONR.
An extended abstract of this paper appears in the Proceedings of the 2007 ACM Workshop on Privacy in the Electronic Society.

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

orously analyzed or even fully stated. This is no surprise, because practical anonymity systems have been deployed and available to study for perhaps a decade, while practical systems for communications confidentiality and/or authenticity have been in use almost as long as there have been electronic communications. It often takes a while for theory and practice to catch up to each other.

Of the many anonymous-communication design proposals (*e.g.* [Chaum 1981; 1988; Reiter and Rubin 1998; Beimel and Dolev 2003; Nambiar and Wright 2006; Corrigan-Gibbs and Ford 2010]), onion routing [Goldschlag et al. 1996] has had notable success in practice. Several implementations have been made [Goldschlag et al. 1996; Syverson et al. 2000; Dingledine et al. 2004], and there was a similar commercial system, Freedom [Goldberg and Shostack 2001]. As of October 2011, Tor [Dingledine et al. 2004], the most recent iteration of the basic design, consists of about 3000 routers, provides a total bandwidth of over 1000 MB/s, and has an estimated total user population of about 500,000 [Loesing et al. 2011]. Because of this popularity, we believe it is important to improve our understanding of the protocol.

Onion routing is a practical anonymity-network scheme with relatively low overhead and latency. Users use a dedicated set of *onion routers* to forward their traffic, obscuring the relationship between themselves and their destinations. To communicate with a destination, a user selects a sequence of onion routers and constructs a *circuit*, or persistant connection, over that sequence. Messages to and from the destination are sent over the circuit. Onion routing provides two-way, connection-based communication and does not require that the destination participate in the anonymity-network protocol. These features make it useful for anonymizing much of the communication that takes place over the Internet today, such as web browsing, chatting, and remote login. Thus, formal analysis and provable anonymity results for onion routing are significant.

As a step toward the overall goal of bridging the gap between provability and practicality in anonymous-communication systems, we have formally modeled and analyzed *relationship anonymity* [Pfitzmann and Hansen 2000; Shmatikov and Wang 2006] in Tor. Although this provides just a small part of the complete understanding of practical anonymity at which our research program is aimed, already it yields nontrivial results that require delicate probabilistic analysis. We hope that this aspect of the work will spur the Theoretical Computer Science community to devote the same level of attention to the rigorous study of anonymity as it has to the rigorous study of confidentiality.

1.1. Summary of Contributions

Black-box abstraction. In the present paper, we treat the network simply as a "black box"[1] to which users connect and through which they communicate with destinations. The abstraction captures the relevant properties of a protocol execution that the adversary can infer from his observations - namely, the observed users, the observed destinations, and the possible connections between the two. In this way, we abstract away from much of the design specific to onion routing so that our results apply both to onion routing and to other low-latency anonymous-communication designs. We express the black-box model within the Universally Composable (UC) security framework [Canetti 2000], which is a standard way to express the function and security properties of cryptographic protocols. We tie our functionality to the guarantees of an actual protocol by showing it reveals as much information about users' communication as the onion routing protocol we formalized [Feigenbaum et al. 2007] in an I/O-automata model.

[1]We note that our use of a "black box" is slightly different than the more common uses in the literature. Black-box access to some cryptographic primitives is commonly used as a starting point to achieve some other desired functionality. Here we show how, for purposes of anonymity analysis, we need only consider a black-box abstraction.

Moreover, we discuss how the functionality can be a protocol within the UC framework itself.

Probabilistic model. Our previous analysis in the I/O-automata model was possibilistic, a notion of anonymity that is simply not sensitive enough. It makes no distinction between communication that is equally likely to be from any one of a hundred senders and communication that came from one sender with probability $.99$ and from each of the other 99 senders with probability $.000101$. An adversary in the real world is likely to have information about which scenarios are more realistic than others. In particular, users' communication patterns are not totally random. When the adversary can determine with high probability, *e.g.*, the sender of a message, that sender is not anonymous in a meaningful way.

Using this intuition, we include a probability measure in our black-box model. In this probability measure, each user chooses a destination according to some probability distribution. We model heterogeneous user behavior by allowing this distribution to be different for different users. We also assume that the users choose their circuits by selecting the routers on it independently and at random.

Bounds on anonymity. We analyze *relationship anonymity* [Pfitzmann and Hansen 2000; Shmatikov and Wang 2006] in our onion routing model. Relationship anonymity is obtained when the adversary cannot identify the destination of a user. The adversary can infer a probability distribution for a user's destination given the adversary's observations. We consider the probability assigned to the correct destination as a measure of anonymity. To be more precise, because this probability depends on the choices of the other users and thus has its own distribution, we will use its expectation as our metric. Moreover, this expectation depends on the other users' destination distributions. If their distributions are very different from that of the given user, the adversary may have an easy time separating out the actions of the user. If they are similar, the user may more effectively hide in the crowd. We provide the following results on a user's anonymity and its dependence on other user behavior:

(1) We show that a standard approximation to our metric provides a lower bound on it (Thm. 3.3).
(2) We show that the worst case for anonymity over other users' behavior is when every other user either always visits the destinations the user is otherwise least likely to visit or always visits his actual destination (Cor. 3.7). The former will be the worst case in most situations.
(3) We give an asymptotic expression for our metric in the worst cases (Thm. 3.6). The limit of this expression in the most common worst case with an adversary controlling a fraction b of the network is equal to the lower bound on the metric when the adversary controls a larger fraction $\sqrt{b}$ of the network. This is significantly worse than the standard analysis suggested, and shows the importance of carefully considering the adversary's knowledge of the system.
(4) We consider anonymity in a more typical set of user distributions in which each user selects a destination from a common Zipfian distribution. Because the users are identical, every user hides well among the others, and we show that, as the user population grows, the anonymity approaches the lower bound (Thm. 4). This shows you may be able to use the standard approximation with accurate results if you are able to make assumptions about user behavior.

1.2. Organization of Paper

We present the details of our black-box model and our anonymity metric in Section 2. In that section we also justify the model by showing how its results provably relate to results in more detailed protocol formalizations ([Feigenbaum et al. 2007; Backes

upon receiving $d \in \Delta$ from $u \in U$

$$x \leftarrow \begin{cases} u & \text{with probability } b \\ \perp & \text{with probability } 1-b \end{cases}$$

$$y \leftarrow \begin{cases} d & \text{with probability } b \\ \perp & \text{with probability } 1-b \end{cases}$$

send (x, y) to the adversary

Fig. 1. Black-box ideal functionality $\mathcal{F}_{OR}$

et al. 2012]). Section 3 presents our results bounding anonymity in our model. Section 4 presents an approximation for anonymity under an additional assumption about typical user behavior. Section 5 describes related work in this area. Finally, we sum up and discuss future research directions in Section 6.

2. TECHNICAL PRELIMINARIES

2.1. Model

We describe our analysis of onion routing in terms of an ideal functionality in the Universal Composability framework [Canetti 2000] We use such a functionality for three reasons: First, it abstracts away the details that aren't relevant to anonymity, second, the UC framework provides the notion of UC emulation, which captures exactly when our analysis applies to a cryptographic protocol, and third, it immediately suggests ways to perform similar analyses of other anonymous-communication protocols that may not strictly provide this functionality.

Let U be the set of users with $|U| = n$. Let Δ be the set of destinations. Let R be the set of onion routers. Let $\mathcal{F}_{OR}$ be the ideal functionality. $\mathcal{F}_{OR}$ takes the set $A \subseteq R$ of compromised routers from the adversary at the beginning of the execution[2]. Let $b = |A|/|R|$. The black-box functionality is given in Figure 2.1. When user u forwards his input from the environment to $\mathcal{F}_{OR}$, the functionality checks to see if it is some $d \in \Delta$. If so, $\mathcal{F}_{OR}$ notifies the adversary of the connection and includes the source with probability b and the destination with probability b.

To analyze the anonymity provided by the ideal functionality, we make two assumptions about the inputs from the environment. First, we assume that the environment selects the destination of user u from a distribution p^u over Δ, where we denote the probability that u chooses d as p^u_d. Second, we assume that the environment sends a destination to each user. Note that these assumptions need not be made when showing that a protocol UC-emulates $\mathcal{F}_{OR}$.

We refer to the combination of the adversary model, the assumptions about the environment, and the ideal functionality as the *black-box model*. Let C be the relevant *configuration* resulting from an execution. C includes a selection of a destination by each user, $C_D : U \to \Delta$, a set of users whose inputs are observed, $C_I : U \to \{0, 1\}$, and a set of users whose outputs are observed, $C_O : U \to \{0, 1\}$. A user's input, output, and destination will be called its *circuit*.

For any configuration, there is a larger set of configurations that are consistent with the outputs that the adversary receives from $\mathcal{F}_{OR}$. We will call two configurations *indistinguishable* if the set of messages (x, y) revealed to the adversary are the same. We use the notation $C \approx \overline{C}$ to indicate that configurations C and $\overline{C}$ are indistinguishable.

Our ideal functionality models anonymous communication over some period of time. It takes as input from each user the identity of a destination. For every such connection

[2]The adversary compromises routers only because a compromised user has no anonymity and is effectively removed from the set of users U for purposes of deanonymization

between a user and destination, the functionality may reveal to the adversary identity of the user, the identity of the destination, or both. Revealing the user corresponds in onion routing to the first router in the circuit being compromised, and revealing the destination corresponds to the last router being compromised. The adversary captured in our model is computationally bounded, controls a fixed set of routers, and can actively attack the protocol. That such an attacker can sometimes learn the source and destination and can link them together is motivated by the results in [Feigenbaum et al. 2007], which we explicitly relate to our ideal functionality in Sec. 2.4. We note that we include only information flow to the adversary in this functionality rather than try to capture the type of communication primitive offered by onion routing because our focus is analyzing anonymity rather than defining a useful anonymous-communication functionality. This model is reminiscent of the general model of anonymous communication used by Kesdogan et al. [2002] in their analysis of an intersection attack. However, we do make a few assumptions that are particularly appropriate for onion routing.

First, the functionality allows the adversary to know whether or not he has directly observed the user. This is valid under the assumption that the initiating client is not located at an onion router itself. This is the case for the vast majority of circuits in Tor and in all significant deployments of onion routing and similar systems to date. We discuss this assumption further in Section 6.

Second, we assume that every user is responsible for exactly one connection in a round. Certainly users can communicate with multiple destinations simultaneously in actual onion-routing systems. However, it seems likely that in practice most users have at most some small (and fixed-bound) number of active connections at any time. To the extent that multiple connections are only slightly more likely to be from the same user than if all connections were independently made and identically distributed, this is a reasonable approximation. This is increasingly true as the overall number of connections grows. To the extent that multiple connections are less likely to be from the same user this is a conservative assumption that gives the adversary as much power to break anonymity as the limited number of user circuits can provide.

Third, the functionality omits the possibility that the adversary observes the user and destination but does not recognize that those observations are part of the same connection. This is another conservative assumption that is motivated by the existence of timing attacks that an active adversary can use to link traffic that it sees at various points along its path through the network [Syverson et al. 2000]. In a timing attack, the adversary observes the timing of the messages going into the onion-routing network and matches them to similar patterns of messages coming out of the onion-routing networks slightly later. Such attacks have been experimentally demonstrated [Øverlier and Syverson 2006; Bauer et al. 2007] and are easy to mount.

Our model captures several different flavors of onion routing (e.g. [Goldschlag et al. 1996; Dingledine et al. 2004; Øverlier and Syverson 2007; Kate et al. 2007]) and possibly some related protocols. Some onion-routing variants, however, do not seem to map well into the abstraction. We discuss this in more detail in Section 5.

Note that our model does not capture several known attacks on anonymity in onion routing. In particular, it does not include attacks exploiting resource interference [Murdoch and Danezis 2005; Murdoch 2006], heterogeneity on network latency [Hopper et al. 2010], correlated destinations between rounds, and identifying patterns of communication [Herrmann et al. 2009]. We do not include such attacks primarily to focus on the most important threats to anonymity, because many of the omitted attacks are attacks on underlying systems rather than on the protocol (e.g., interference) or have limited effectiveness or are mitigated by improvements to the protocol. Also, we see

the analysis of our simplified model as a first step in establishing rigorous guarantees of anonymity in increasingly realistic models.

2.2. Probabilistic Anonymity

A user performs an action anonymously in a possibilistic sense if there is an indistinguishable configuration in which the user does not perform the action. For example, under this definition a user with observed output but unobserved input sends that output anonymously if there exists another user with unobserved input. The probability measure we have added to configurations allows us to incorporate the degree of certainty that the adversary has about the subject of an action. After making observations in the actual configuration, the adversary can infer a conditional probability distribution on configurations. There are several candidates in the literature for assessing an anonymity metric from this distribution. The probabilistic anonymity metric that we use is the posterior probability of the correct subject. The lower this is, the more anonymous we consider the user. In part, we use this metric because it is simple. Also, any statements about entropy and maximum probability metrics only make loose guarantees about the probability assigned to the actual subject, a quantity that clearly seems important to the individual users.

Observe that this choice assumes that the adversary has perfect prior information about the system. He may not actually know the underlying probability measure, however. In particular, it doesn't seem likely that the adversary would know how every user selects destinations. In our analysis, we take a worst-case view and assume that the adversary knows the distributions exactly. Also, over time he might learn a good approximation of user behavior via the long-term intersection attack [Danezis and Serjantov 2004]. In this case, it may seem as though anonymity has been essentially lost anyway. However, even when the adversary knows how a user generally behaves, the anonymity network may make it hard for him to determine who is responsible for any specific action, and the anonymity of a specific action is what we are interested in.

2.3. Relationship Anonymity

We analyze the *relationship anonymity* of users and destinations in our model, that is, how well the adversary can determine if a user and destination have communicated. Our metric for the relationship anonymity of user u and destination d is the posterior probability ψ that u chooses d as his destination. We study ψ directly, although the *anonymity* of a user's communication with a destination is $1 - \psi$.

Using the posterior probability makes sense in this context because it focuses on the information that users are trying to hide—their actual destinations—without being affected by information the adversary learns about other destinations. Onion routing does leak information, and using a metric such as the entropy of the posterior distribution or the statistical distance from the prior may not give a good idea of how well the adversary can correctly guess the user's behavior. Designers may wish to know how well a system protects communications on average or overall. But it is also important for a user to be able to assess how secure he can expect a particular communication to be in order to decide whether to create it or not. This is the question we address. Moreover, the metric is relatively simple to analyze. Furthermore, to the extent that the user may not know how he fits in and thus wishes to know the worst risk for any user, that is just a lower bound on our metric.

The relationship anonymity of u and d varies with the destination choices of the other users and the observations of the adversary. If, for example, u's output is observed, and the inputs of all other users are observed, then the adversary knows u's destination with probability 1. Because we want to examine the relationship anonymity of u conditioned only on his destination, we end up with a distribution

on the anonymity metric. We look at the expectation of this distribution. Moreover, because this distribution depends on the destination distributions of all of the users, we continue by finding the worst-case expectation in the limit for a given user and destination and then examine the expectation in a more likely situation.

2.4. Emulating the Ideal Functionality

The anonymity analysis of the ideal functionality $\mathcal{F}_{OR}$ that we perform in Sections 3 and 4 is meaningful to the extent that $\mathcal{F}_{OR}$ captures the information that an adversary can obtain by interacting with onion-routing protocols. We justify the functionality primarily by showing that it provides the same information about the source of a given connection as does onion-routing as captured in our previous formalization [Feigenbaum et al. 2007]. Furthermore, we describe separate work showing that $\mathcal{F}_{OR}$ can be UC-emulated by an onion-routing protocol.

Relationship to I/O-automata model. We have formalized onion routing using an I/O-automata model [Lynch 1996] and an idealization of the cryptographic properties of the protocol [Feigenbaum et al. 2007]. Their analysis identifies the user states that are information-theoretically indistinguishable. The black-box model we provide herein is a valid abstraction of that formalization because, under some reasonable probability measures on executions, it preserves the relationship-anonymity properties.

The I/O-automata model includes a set of users U, a set of routers R, an adversary $A \subseteq R$, and a set of destinations Δ, where we take the final router in the I/O-automata model to be the destination and assume that it is uncompromised. A configuration C in the I/O-automata model is a mapping from each user $u \in U$ to a circuit $(r_1^u, \ldots, r_l^u) \in R^l$, a destination $d^u \in \Delta$, and a circuit identifier $n^u \in \mathbb{N}_+$. An *execution* is a sequence of I/O-automaton states and actions, which must be consistent with the configuration.

Let users in the I/O-automata model choose the other routers in their circuits uniformly at random and choose the destination according to user-specific distributions. Given these circuits and a set of adversary automata, we have previously identifed [Feigenbaum et al. 2007] an equivalence class of circuit and destination choices with respect to which, for every pair of configurations in the class, a bijection exists between their executions such that paired executions are indistinguishable. Let the indistinguishable executions thus paired have the same probability, conditional on their configuration.

Given this measure, the black-box model that abstracts the I/O-automata model has the same user set U, the same destination set Δ, an adversary parameter of $b = |A|/|R|$, and the same destination distributions. The following theorem shows that each posterior distribution on the destinations of users has the same probability under both the I/O-automata model and its black-box model. Let E be a random I/O-automata execution. Let X^a be a random I/O-automata configuration (X^a can be viewed as a function mapping a random execution to its configuration). Let X^b be a random black-box configuration. Let $\psi_1(u, d, E)$ be the posterior probability that u visited d in the I/O-automata model, *i.e.*, the conditional given that the execution is indistinguishable from E. Let $\psi_2(u, d, X^b)$ be the posterior probability that u visited d in the black-box model, *i.e.*, the conditional distribution given that the configuration is indistinguishable from X^b. Let $\psi_0(u, d)$ be a distribution over destinations d for every u.

THEOREM 2.1.

$$Pr[\forall_{u \in U, d \in \Delta} \psi_1(u, d, E) = \psi_0(u, d)] = Pr[\forall_{u \in U, d \in \Delta} \psi_2(u, d, X^b) = \psi_0(u, d)]$$

PROOF. Let ϕ be the map from I/O-automata configurations to black-box configurations such that

(1) $\phi(C)_D(u) = d^u$

(2) $\phi(C)_I(u) = \begin{cases} 1 & \text{if } r_1 \in A \\ 0 & \text{otherwise} \end{cases}$

(3) $\phi(C)_O(u) = \begin{cases} 1 & \text{if } r_l \in A \\ 0 & \text{otherwise} \end{cases}$.

ϕ essentially "quotients out" the specific router choices of each user, retaining the compromised status of the first and last routers as well as the destination. It allows us to relate the posterior ψ_1 in the I/O-automata model to the ψ_2 in the black-box model.

Let C_1^a be any I/O-automata configuration. Given any execution e of C_1^a, the adversary's posterior probability on configurations is

$$\frac{Pr[X^a = C_2^a]}{\sum_{C^a \approx C_1^a} Pr[X^a = C^a]}$$

if $C_2^a \approx C_1^a$ and 0 otherwise, because we set equal the probability of two executions that are paired with each other in the bijection on executions constructed in Feigenbaum et al. [2007]. Because the configurations determine which destination each user visits, the distribution $\psi_1(u, d, e)$ can be determined from the posterior distribution on configurations. Notice that this distribution only puts positive probability on the set $\mathcal{C}^a$ of configurations that are indistinguishable from C_1^a.

The posterior distribution on I/O-automata configurations induces a posterior distribution on black-box configurations via ϕ. ϕ preserves the destination of each user, and so the distribution $\psi_1(u, d, e)$ can be determined from this distribution on black-box configurations. Notice that this distribution only puts positive probability on the set of black-box configurations $\phi(\mathcal{C}^a)$ that are mapped to by I/O-automata configurations in $\mathcal{C}^a$.

To understand the set $\phi(\mathcal{C}^a)$ and its posterior distribution given e, consider the equivalence class $\mathcal{C}^b$ of the configuration $\phi(C_1^a)$. Let S be those configurations in $\mathcal{C}^a$ that differ from C_1^a only in the destinations and the permutation of users. From Theorems 1 and 2 in [Feigenbaum et al. 2007], it follows that ϕ is a bijection between S and $\mathcal{C}^b$. The posterior probability of each $C_2^a \in S$ is proportional to $Pr[X^b = \phi(C_2^a)]$ because the prior probability of C_2^a is $Pr[X^b = \phi(C_2^a)]$ multiplied by the probability of selecting its given routers (which are the same for all $s \in S$) given that $\phi(X^a) = \phi(C_2^a)$. Moreover, all of the other configurations in C^a are reached by changing the unobserved routers of one of the configurations in S. ϕ is invariant under such a change. Also, the posterior probability is invariant under such a change because the routers are chosen independently and uniformly at random. Furthermore, the number of I/O-automata configurations that are reached by such a change from some $s \in S$ is the same for all s. Therefore, the posterior probability $Pr[\phi(X^a) = C^b | e]$ is proportional to $Pr[X^b = C^b]$ for $C^b \in \mathcal{C}^b$, and is zero otherwise. Therefore, $\psi_1(u, d, e) = \psi_2(u, d, \phi(C_1^a))$.

By this equality, the probability that a random execution E results in a given posterior $\psi_0(u, d)$ is equal to the probability that the I/O-automata configuration X^a maps under ϕ to a black-box configuration $\phi(X^a) = C^b$ such that $\psi_2(u, d, C^b) = \psi_0(u, d)$. The probability $Pr[\phi(X^a) = C^b]$ is equal to $Pr[X^b = C^b]$ because the probability of first-router compromise and the probability of an input being observed are both b, last-router compromise and an output being observed are both independent events with probability b, and user destinations are chosen independently in both models and follow the same distributions. Therefore,

$$Pr[\forall_{u \in U, d \in \Delta} \psi_1(u, d, E) = \psi_0(u, d)] = Pr[\forall_{u \in U, d \in \Delta} \psi_2(u, d, X) = \psi_0(u, d)].$$

□

UC emulation. Expressing our black-box model within the UC framework allows it to be compared to protocols expressed within the same framework. In particular, if a protocol can be shown to UC-emulate $\mathcal{F}_{OR}$, then, making only common cryptographic assumptions, the adversary can make only negligibly better guesses about users' communication when interacting with that protocol than he can interacting with the functionality. Backes et al. [2012] show that such an argument can indeed be made. They give an onion-routing protocol, show that it UC-emulates our black-box functionality, and use this result to apply our results about anonymity to their system. For UC emulation, it must be shown that an adversary cannot determine whether he is interacting with the actual protocol or with a simulator that is itself only interacting with the ideal functionality. Emulation of $\mathcal{F}_{OR}$ by an onion-routing protocol is achieved with a simulator that makes all of the protocol decisions left undetermined by the interface. That is, given partial information about a new connection from $\mathcal{F}_{OR}$, the simulator chooses an onion-routing circuit consistent with that information and simulates the construction of that circuit.

3. EXPECTED ANONYMITY

Let the set $\mathcal{C}$ of all configurations be the sample space and X be a random configuration. Let Ψ be the posterior probability of the event that u chooses d as a destination, that is, $\Psi(C) = Pr[X_D(u) = d | X \approx C]$. Ψ is our metric for the relationship anonymity of u and d.

Let $\mathbb{N}^{\Delta}$ represent the set of multisets over Δ. Let $\rho(\Delta^0)$ be the maximum number of orderings of $\Delta^0 \in \mathbb{N}^{\Delta}$ such that the same destination is in any given location in every ordering:

$$\rho(\Delta^0) = \prod_{\delta \in \Delta} |\{\delta \in \Delta^0\}|!$$

Let $\Pi(A, B)$ be the set of all injective maps $A \to B$. The following theorem gives an exact expression for the conditional expectation of Ψ in terms of the underlying parameters U, Δ, p, and b:

THEOREM 3.1.

$$\begin{aligned} E[\Psi | X_D(u) = d] = b(1-b)p_d^u + b^2 + \\ \sum_{S \subseteq U : u \in S} \sum_{\Delta^0 \in \mathbb{N}^{\Delta} : |\Delta^0| \leq S} b^{n - |S| + |\Delta^0|} (1-b)^{2|S| - |\Delta^0|} \cdot \\ \left(\sum_{T \subseteq S - u : |T| = |\Delta^0| - 1} \sum_{\pi \in \Pi(T + u, \Delta^0) : \pi(u) = d} p_d^u \prod_{v \in T} p_{\pi(v)}^v \right. \\ \left. + \sum_{T \subseteq S - u : |T| = |\Delta^0|} \sum_{\pi \in \Pi(T, \Delta^0)} p_d^u \prod_{v \in T} p_{\pi(v)}^v \right)^2 \cdot \\ [\rho(\Delta^0)]^{-1} (p_d^u)^{-1} \left(\sum_{T \subseteq S : |T| = |\Delta^0|} \sum_{\pi \in \Pi(T, \Delta^0)} \prod_{v \in T} p_{\pi(v)}^v \right)^{-1} \end{aligned} \quad (1)$$

PROOF. At a high level, the conditional expectation of Ψ can be expressed as:

$$\mathbf{E}[\Psi | X_D(u) = d] = \sum_{C \in \mathcal{C}} \mathbf{Pr}[X = C | X_D(u) = d] \Psi(C).$$

We calculate Ψ for a configuration C by finding the relative weight of indistinguishable configurations in which u selects d. The adversary observes some subset of the circuits. If we match the users to circuits in some way that sends users with observed inputs to their own circuits, the result is an indistinguishable configuration. Similarly, we can match circuits to destinations in any way that sends circuits on which the output has been observed to their actual destination in C.

The value of $\Psi(C)$ is especially simple if u's input has been observed. If the output has not also been observed, then $\Psi(C) = p_d^u$. If the output has also been observed, then $\Psi(C) = 1$.

For the case in which u's input has not been observed, we have to take into account the destinations of and observations on the other users. Let $S \subseteq U$ be the set of users s such that $C_I(s) = 0$. Note that $u \in S$. Let Δ^0 be the multiset of the destinations of circuits in C on which the input has not been observed, but the output has.

Let $f_0(S, \Delta^0)$ be the probability that in a random configuration the set of unobserved inputs is S and the set of observed destinations with no corresponding observed input is Δ^0:

$$f_0(S, \Delta^0) = b^{n-|S|+|\Delta^0|}(1-b)^{2|S|-|\Delta^0|}[\rho(\Delta^0)]^{-1} \sum_{T \subseteq S : |T| = |\Delta^0|} \sum_{\pi \in \Pi(T, \Delta^0)} \prod_{v \in T} p^v_{\pi(v)}.$$

Let $f_1(S, \Delta^0)$ be the probability that in a random configuration the set of unobserved inputs is S, the set of observed destinations with no corresponding observed input is Δ^0, the output of u is observed, and the destination of u is d:

$$f_1(S, \Delta^0) = b^{n-|S|+|\Delta^0|}(1-b)^{2|S|-|\Delta^0|}[\rho(\Delta^0)]^{-1} p_d^u \cdot \sum_{T \subseteq S-u : |T| = |\Delta^0|-1} \sum_{\pi \in \Pi(T+u, \Delta^0) : \pi(u) = d} \prod_{v \in T} p^v_{\pi(v)}.$$

Let $f_2(S, \Delta^0)$ be the probability that in a random configuration the set of unobserved inputs is S, the set of observed destinations with no corresponding observed input is Δ^0, the output of u is unobserved, and the destination of u is d:

$$f_2(S, \Delta^0) = b^{n-|S|+|\Delta^0|}(1-b)^{2|S|-|\Delta^0|}[\rho(\Delta^0)]^{-1} p_d^u \cdot \sum_{T \subseteq S-u : |T| = |\Delta^0|} \sum_{\pi \in \Pi(T, \Delta^0)} \prod_{v \in T} p^v_{\pi(v)}.$$

Now we can express the posterior probability $\Psi(C)$ as:

$$\Psi(C) = \frac{f_1(S, \Delta^0) + f_2(S, \Delta^0)}{f_0(S, \Delta^0)}. \tag{2}$$

The expectation of Ψ is a sum of the above posterior probabilities weighted by their probability. The probability that the input of u has been observed but the output hasn't is $b(1-b)$. The probability that both the input and output of u have been observed is b^2. These cases are represented by the first two terms in Equation 1.

When the input of u has not been observed, we have an expression of the posterior in terms of sets S and Δ^0. The numerator ($f_1 + f_2$) of Equation 2 itself actually sums the weight of every configuration that is consistent with S, Δ^0, and the fact that the destination of u is d. However, we must divide by p_d^u, because we condition on the event $\{X_D(u) = d\}$.

These observations give us the remaining terms in Equation 1. □

3.1. Simple approximation of conditional expectation

The expression for the conditional expectation of Ψ in Equation 1 is difficult to interpret. It would be nice if we could find a simple approximation. The probabilistic analysis in Syverson et al. [2000] proposes just such a simplification by reducing it to only two cases: i) the adversary observes the user's input and output and therefore identifies his destination, and ii) the adversary doesn't observe these and cannot improve his *a priori* knowledge. The corresponding simplified expression for the expection is:

$$E[\Psi|X_D(u) = d] \approx b^2 + (1 - b^2)p_d^u. \tag{3}$$

This is a reasonable approximation if the final summation in Equation 1 is about $(1 - b)p_d^u$. This summation counts the case in which u's input is not observed, and to achieve a good approximation the adversary must experience no significant advantage or disadvantage from comparing the users with unobserved inputs (S) with the discovered destinations (Δ^0).

The quantity $(1 - b)p_d^u$ does provide a lower bound on the final summation. It may seem obvious that considering the destinations in Δ^0 can only improve the accuracy of adversary's prior guess about u's destination. However, in some situations the posterior probability for the correct destination may actually be smaller than the prior probability. This may happen, for example, when some user v, $v \neq u$, communicates with a destination e, $e \neq d$, and only u is *a priori* likely to communicate with e. If the adversary observes the communication to e, it may infer that it is likely that u was responsible and therefore didn't choose d.

It is true, however, that in expectation this probability can only increase. Therefore Equation 3 provides a lower bound on the anonymity metric.

The proof of this fact relies on the following lemma. Let $\mathcal{E}$ be an event in some finite sample space Ω. Let $\mathcal{A}_1, \ldots, \mathcal{A}_n$ be a set of disjoint events such that $\mathcal{E} \subseteq \bigcup_i \mathcal{A}_i$, and let $\mathcal{A}^j = \bigcup_{i=1}^{j} \mathcal{A}_i$. Let $\mathcal{E}_i = \mathcal{E} \cap \mathcal{A}_i$. Finally, let $Y(\omega) = \sum_i 1_{\mathcal{E}_i}(\omega) Pr[\mathcal{E}_i]/Pr[\mathcal{A}_i]$ (where $1_{\mathcal{E}_i}$ is the indicator function for $\mathcal{E}_i$). $Y(\omega)$ is thus the conditional probability $Pr[\mathcal{E}|\mathcal{A}_i]$, where $\omega \in \mathcal{E}_i$.

LEMMA 3.2. $Pr[\mathcal{E}|\mathcal{A}^n] \leq E[Y|\mathcal{E}]$

PROOF.

$$\begin{aligned}
Pr[\mathcal{E}|\mathcal{A}^n] &= \frac{Pr[\mathcal{E}]}{Pr[\mathcal{A}^n]} \\
&= \frac{\left(\sum_i \frac{Pr[\mathcal{E}_i]\sqrt{Pr[\mathcal{A}_i]}}{\sqrt{Pr[\mathcal{A}_i]}}\right)^2}{Pr[\mathcal{A}^n]Pr[\mathcal{E}]} && \text{by a simple rewriting} \\
&\leq \frac{\left(\sqrt{\sum_i \frac{(Pr[\mathcal{E}_i])^2}{Pr[\mathcal{A}_i]}}\sqrt{\sum_i Pr[\mathcal{A}_i]}\right)^2}{Pr[\mathcal{A}^n]Pr[\mathcal{E}]} && \text{by the Cauchy-Schwartz inequality} \\
&= \sum_i \frac{(Pr[\mathcal{E}_i])^2}{Pr[\mathcal{A}_i]Pr[\mathcal{E}]} \\
&= E[Y|\mathcal{E}]
\end{aligned}$$

□

THEOREM 3.3. $E[\Psi|X_D(u) = d] \geq b^2 + (1 - b^2)p_d^u$

PROOF. As described in the proof of Theorem 3.1:

$$E[\Psi|X_D(u) = d] = b^2 + b(1 - b)p_d^u + (1 - b)E[\Psi|X_D(u) = d \wedge X_I(u) = 0].$$

To apply Lemma 3.2, take the set of configurations $\mathcal{C}$ to be the sample space Ω. Take $\{X_D(u) = d\}$ to be the event $\mathcal{E}$. Take the indistinguishability equivalence relation to be the sets $\mathcal{A}_i$. Finally, take Ψ to be Y. Then the lemma shows that $E[\Psi|X_D(u) = d \wedge X_I(u) = 0] \geq p_d^u$. □

3.2. Worst-case Anonymity

To examine the accuracy of our approximation, we look at how large the final summation in Equation 1 can get as the users' destination distributions vary. Because this is the only term that varies with the other user distributions, this will also provide a worst-case guarantee on expected anonymity metric. Our results will show that, in the limit as the number of users grows, the worst case can occur when the users other than u act as differently from u as possible by always visiting the destination u is otherwise least likely to visit. Less obviously, we show that the limiting maximum can also occur when the users other than u always visit d. This happens because it makes the adversary observe destination d often, causing him to suspect that u chose d. Our results also show that the worst-case expectation is about $b + (1-b)p_d^u$, which is significantly worse than the simple approximation above.

As the first step in finding the maximum of Equation 1 over $(p^v)_{v \neq u}$, we observe that it is obtained when every user $v \neq u$ chooses only one destination d_v, *i.e.* $p^v_{d_v} = 1$ for some $d_v \in \Delta$.

LEMMA 3.4. *A maximum of $E[\Psi | X_D(u) = d]$ over $(p^v)_{v \neq u}$ must occur when, for all $v \neq u$, there exists some $d_v \in \Delta$ such that $p^v_{d_v} = 1$.*

PROOF. Take some user $v \neq u$ and two destinations $e, f \in \Delta$. Assign arbitrary probabilities in p^v to all destinations except for f, and let $\zeta = 1 - \sum_{\delta \neq e,f} p^v_\delta$. Then $p^v_f = \zeta - p^v_e$. Consider $E[\Psi | X_D(u) = d]$ as a function of p^v_e. The terms t_i of Equation 1 that correspond to any fixed S and Δ^0 are of the following general form, where $c^i_1, c^i_2, c^i_3, c^i_4, c^i_5, c^i_6 \geq 0$:

$$t_i = \frac{(c^i_1 p^v_e + c^i_2(\zeta - p^v_e) + c^i_3)^2}{c^i_4 p^v_e + c^i_5(\zeta - p^v_e) + c^i_6}.$$

This is a convex function of p^v_e:

$$D^2_{p^v_e} t_i = \frac{2(c^i_3(c^i_4 - c^i_5) + c^i_2(c^i_4\zeta + c^i_6) - c^i_1(c^i_5\zeta + c^i_6))^2}{(c^i_5(\zeta - p^v_e) + c^i_4 p^v_e + c^i_6)^3} \geq 0.$$

The leading two terms of $E[\Psi | X_D(u) = d]$ are constant in p^v, and the sum of convex functions is a convex function, so $E[\Psi | X_D(u) = d]$ is convex in p^v_e. Therefore, a maximum of $E[\Psi | X_D(u) = d]$ must occur when $p^v_e \in \{0, 1\}$. □

Order the destinations $d = d_1, \ldots, d_{|\Delta|}$ such that $p^u_{d_i} \geq p^u_{d_{i+1}}$ for $i > 1$. The following lemma shows that we can further restrict ourselves to distribution vectors in which, for every user except u, the user either always chooses d or always chooses $d_{|\Delta|}$.

LEMMA 3.5. *A maximum of $E[\Psi | X_D(u) = d]$ must occur when, for all users v, either $p^v_{d_1} = 1$ or $p^v_{d_{|\Delta|}} = 1$.*

PROOF. Assume, following Lemma 3.4, that $(p^v)_{v \neq u}$ is an extreme point of the set of possible distribution vectors.

Equation 1 groups configurations first by the set S with unobserved inputs and second by the observed destinations Δ^0. Instead, group configurations first by S and second by the set $T \subseteq S$ with observed outputs. Because every user except u chooses a destination deterministically, Ψ only depends on the sets S and T. Let $\Psi_1(S, T)$ be this value.

$$E[\Psi | X_D(u) = d] = b(1-b)p^u_d + b^2 + \sum_{S : u \in S} \sum_{T : T \subseteq S} b^{n - |S| + |T|}(1-b)^{2|S| - |T|} \Psi_1(S, T). \quad (4)$$

Select two destinations $d_i, d_j, 1 < i < j$. We break up the sum in Equation 4 and show that, for every piece, the sum can only be increased by changing $(p^v)_v$ so that any user that always chooses d_i always chooses d_j instead.

Fix $S \subseteq U$ such that $u \in S$. Let $S_i, S_j \subseteq S$ be such that $p^s_{d_i} = 1$ if and only if $s \in S_i$, and $p^s_{d_j} = 1$ if and only if $s \in S_j$. Fix $T' \subseteq S \backslash S_i \backslash S_j$ and some $t \geq |T'|$.

Let $f(S, T')$ be the sum of terms in Equation 4 that are indexed by S and some T such that $|T| = t$ and $T \supseteq T'$. To calculate $f(S, T')$, group its terms by the number t_{d_i} of users v in T such that $X_D(v) = d_i$. Let t_e be the number for these terms of users v in T' such that $X_D(v) = e$, $e \in \Delta \backslash \{d_i, d_j\}$. The number t_{d_j} of users v such that $X_D(v) = d_j$ for these terms is then $t - \sum_{e \in \Delta - d_j} t_e$. Let s_e be the number of users v in $S - u$ such that $X_D(v) = e$. The number of terms in $f(S, T')$ with a given t_{d_i} is then

$$\binom{s_{d_i}}{t_{d_i}} \binom{s_{d_j}}{t_{d_j}}.$$

For each of these terms, Ψ_1 is the same. To calculate it, let f_δ be the number of configurations that yield the given S and $(t_e)_{e \in \Delta}$ and are such that u's output is observed with destination δ:

$$f_\delta(t_{d_i}) = \binom{s_\delta}{t_\delta - 1} \prod_{e \in \Delta - \delta} \binom{s_e}{t_e},$$

and let f_0 be the number of configurations that yield the same S and $(t_e)_{e \in \Delta}$ and are such that u's output is unobserved:

$$f_0(t_{d_i}) = \prod_{e \in \Delta} \binom{s_e}{t_e}.$$

Then the posterior probability given S and $(t_e)_{e \in \Delta}$ is

$$\frac{p^u_d \left(f_d(t_{d_i}) + f_0(t_{d_i})\right)}{\sum_{\delta \in \Delta} p^u_\delta f_\delta(t_{d_i}) + f_0(t_{d_i})}.$$

Therefore, letting $m = t - \sum_{e \in \Delta \backslash \{d_i, d_j\}} t_e$,

$$f(S, T') = b^{n-|S|+t}(1-b)^{2|S|-t} \sum_{t_{d_i}=0}^{m} \binom{s_{d_i}}{t_{d_i}} \binom{s_{d_j}}{m - t_{d_i}} \frac{p^u_d \left(f_d(t_{d_i}) + f_0(t_{d_i})\right)}{\sum_{\delta \in \Delta} p^u_\delta f_\delta(t_{d_i}) + f_0(t_{d_i})}.$$

The binomial coefficients of f_δ and f_0 in the numerator and denominator largely cancel, and the whole expression can be simplified to

$$f(S, T') = \alpha \sum_{t_{d_i}=0}^{m} \binom{s_{d_i}}{t_{d_i}} \binom{s_{d_j}}{m - t_{d_i}} \frac{(s_{d_i} + 1 - t_{d_i})(s_{d_j} + 1 - m + t_{d_i})}{\left(\begin{array}{l} p^u_{d_i}(s_{d_i} + 1)(s_{d_j} + 1 - m + t_{d_i}) + \\ p^u_{d_j}(s_{d_j} + 1)(s_{d_i} + 1 - t_{d_i}) + \\ (s_{d_i} + 1 - t_{d_i})(s_{d_j} + 1 - m + t_{d_i})\beta \end{array}\right)}$$

for some $\alpha, \beta \geq 0$.

This can be seen as the weighted convolution of binomial coefficients. Unfortunately, there is no obvious way to simplify the expression any further to find the maximum as we trade off s_{d_i} and s_{d_j}. There is a closed-form sum if the coefficient of the binomial product is a fixed-degree polynomial, however. Looking at the coefficient, we can see

that it is concave.

$$c_{t_{d_i}} = \frac{(s_{d_i}+1-t_{d_i})(s_{d_j}+1-m+t_{d_i})}{p^u_{d_i}(s_{d_i}+1)(s_{d_j}+1-m+t_{d_i})+p^u_{d_j}(s_{d_j}+1)(s_{d_i}+1-t_{d_i})+(s_{d_i}+1-t_{d_i})(s_{d_j}+1-m+t_{d_i})\beta}.$$

$$\begin{aligned} D^2_{t_{d_i}} c_{t_{d_i}} &= -\frac{\left(\begin{array}{l} 2((s_{d_i}+1)(s_{d_j}+1)(2+s_{d_i}+s_{d_j}-m)^2 p^u_{d_i} p^u_{d_j} + \\ b((s_{d_i}+1)(s_{d_j}+1+t_{d_i}-m)^3 p^u_{d_i} + (s_{d_j}+1)(s_{d_i}+1-t_{d_i})^3 p^u_{d_j})) \end{array}\right)}{((s_{d_j}+1+t_{d_i}-m)(b(s_{d_i}+1-t_{d_i})+p^u_{d_i}(s_{d_i}+1))+(s_{d_j}+1)(s_{d_i}+1-t_{d_i})p^u_{d_j})^3} \\ &\leq 0. \end{aligned}$$

We can use this fact to bound the sum above by replacing $c_{t_{d_i}}$ with a line tangent at some point i_0. Call this approximation $\tilde{f}$. Holding $s_{d_i}+s_{d_j}$ constant, this approximation is in fact equal at $s_{d_i} = 0$ because the sum has only one term. Then, if $s_{d_i} = 0$ still maximizes the sum, the theorem is proved. Let $c'_{i_0} = D_{t_{d_i}} c_{t_{d_i}}\big|_{t_{d_i}=i_0}$.

$$\begin{aligned} f(S,T') &\leq \sum_{t_{d_i}=0}^{m} \binom{s_{d_i}}{t_{d_i}} \binom{s_{d_j}}{m-t_{d_i}} (c'_{i_0}(t_{d_i}-i_0)+c_{i_0}) \\ &= \binom{s_{d_i}+s_{d_j}}{m} \left(c_{i_0} + c'_{i_0}\frac{m\cdot s_{d_i}}{s_{d_i}+s_{d_j}} - c'_{i_0} i_0\right) \\ &= \tilde{f}(S,T'). \end{aligned}$$

The linear approximation will be done around the point $i_0 = m \cdot s_{d_i}/(s_{d_i}+s_{d_j})$. This results in a simple form for the resulting approximation, and also the mass of the product of binomial coefficients concentrates around this point. Set $\nu = s_{d_i}+s_{d_j}$ to examine the tradeoff between s_{d_i} and s_{d_j}.

$$\begin{aligned} \tilde{f}(S,T') &= \binom{\nu}{m}\left(c_{\frac{m\cdot s_{d_i}}{\nu}}\right) \\ &= \binom{\nu}{m} \frac{((\nu-s_{d_i})(\nu-m)+\nu)((s_{d_i}+1)\nu - m\cdot s_{d_i})}{\left(\begin{array}{l} p^u_{d_i}\nu(s_{d_i}+1)((\nu-s_{d_i})(\nu-m)+\nu)+ \\ p^u_{d_j}\nu(\nu-s_{d_i}+1)(\nu+s_{d_i}(\nu-m))+ \\ \beta((s_{d_i}+1)\nu - m\cdot s_{d_i})((\nu-s_{d_i})(\nu-m)+\nu) \end{array}\right)}. \end{aligned}$$

Lemma A.1 in the Appendix shows that $\tilde{f}$ is convex in s_{d_i}. Thus, the maximum of $\tilde{f}$ must exist at $s_{d_i} = 0$ or $s_{d_i} = \nu$. Observe that when $s_{d_i} = 0$,

$$\tilde{f} = \binom{\nu}{m}\frac{1-m+\nu}{p_{d_j}(1+\nu)+\beta(1-m+\nu)+p_{d_i}(1-m+\nu)}$$

and when $s_{d_i} = \nu$

$$\tilde{f} = \binom{\nu}{m}\frac{1-m+\nu}{p_{d_j}(1-m+\nu)+\beta(1-m+\nu)+p_{d_i}(1+\nu)}.$$

Therefore, because $p_{d_i} \geq p_{d_j}$, $\tilde{f}$ is larger when $s_{d_i} = 0$. As stated, this implies that f itself is maximized when $s_{d_i} = 0$.

□

Therefore, in looking for a maximum we can assume that every user except u either always visits d or always visits $d_{|\Delta|}$. To examine how anonymity varies with the number of users in each category, we derive an asymptotic estimate for large n. A focus

on large n is reasonable because anonymity networks, and onion routing in particular, are understood to have the best chance at providing anonymity when they have many users. Furthermore, Tor is currently used by an estimated 500,000 people.

Let $\alpha = \{v \neq u : p_d^v = 1\}/(n-1)$ be the fraction of users that always visit d. Theorem 3.6 gives an asymptotic estimate for the expected posterior probability given a constant α. It shows that, in the limit, the maximum expected posterior probability is obtained when all users but u always visit d or when they always visit $d_{|\Delta|}$.

THEOREM 3.6. *Assume that, for all $v \neq u$, either $p_d^v = 1$ or $p_{d_{|\Delta|}}^v = 1$. Then, if $\alpha = 0$,*

$$E[\Psi|X_D(u) = d] = b(1-b)p_d^u + b^2 + (1-b)\left(b + \frac{(1-b)^2 p_d^u}{1-b+p_{d_{|\Delta|}}^u b}\right) + O\left(\sqrt{\frac{\log(n)}{n}}\right),$$

if $0 < \alpha < 1$

$$E[\Psi|X_D(u) = d] = b(1-b)p_d^u + b^2 + (1-b)\frac{p_d^u}{1-b+p_d^u b + p_{d_{|\Delta|}}^u b} + O\left(\sqrt{\frac{\log(n)}{n}}\right),$$

and, if $\alpha = 1$,

$$E[\Psi|X_D(u) = d] = b(1-b)p_d^u + b^2 + (1-b)\frac{p_d^u}{1-b+p_d^u b} + O\left(\sqrt{\frac{\log(n)}{n}}\right).$$

PROOF. Let $n_e = \alpha(n-1)$ and $n_f = (1-\alpha)(n-1)$. The expected posterior probability can be given in the following variation on Equation 4:

$$\begin{aligned} E[\Psi|X_D(u) = d] = b(1-b)p_d^u + b^2 + (1-b)\cdot \\ \sum_{e=0}^{n_e}\binom{n_e}{e}(1-b)^e b^{n_e-e}\sum_{f=0}^{n_f}\binom{n_f}{f}(1-b)^f b^{n_f-f}\cdot \\ \sum_{j=0}^{f}\binom{f}{j}b^j(1-b)^{f-j}\sum_{k=0}^{e}\binom{e}{k}b^k(1-b)^{e-k}\cdot \\ \left[b\Psi_2(e,f,j,k+1) + (1-b)\Psi_2(e,f,j,k)\right]. \end{aligned} \tag{5}$$

Here $\Psi_2(e,f,j,k)$ is the value of Ψ when the users with unobserved inputs consist of u, e users $v \neq u$ with $p_d^v = 1$, and f users $v \neq u$ with $p_{d_{|\Delta|}}^v = 1$; and the users with unobserved inputs and observed outputs consist of k users v with $X_D(v) = d$ and j users v with $X_D(v) = d_{|\Delta|}$. Given such a configuration, the number of indistinguishable configurations in which u has observed destination d is $\binom{e}{k-1}\binom{f}{j}$, the number of indistinguishable configurations in which u has observed destination $d_{|\Delta|}$ is $\binom{e}{k}\binom{f}{j-1}$, and the number of indistinguishable configuration in which u has an unobserved destination is $\binom{e}{k}\binom{f}{j}$. Thus, we can express Ψ_2 as

$$\Psi_2(e,f,j,k) = \frac{p_d^u\binom{e}{k-1}\binom{f}{j} + p_d^u\binom{e}{k}\binom{f}{j}}{p_d^u\binom{e}{k-1}\binom{f}{j} + p_{d_{|\Delta|}}^u\binom{e}{k}\binom{f}{j-1} + \binom{e}{k}\binom{f}{j}}.$$

The binomial coefficients largely cancel, and so we can simplify this equation to

$$\Psi_2(e,f,j,k) = \frac{p_d^u(e+1)(f-j+1)}{p_d^u k(f-j+1) + p_{d_{|\Delta|}}^u j(e-k+1) + (e-k+1)(f-j+1)}.$$

We observe that j and k are binomially distributed. Therefore, by the Chernoff bound, they concentrate around their means as e and f grow. Let $\mu_1 = fb$ be the mean of j and $\mu_2 = eb$ be the mean of k. We can approximate the tails of the sums over j and k in Equation 5 and sum only over the central terms:

$$\begin{aligned}
E[\Psi|X_D(u) = d] = b(1-b)p_d^u + b^2 + \\
(1-b)\sum_{e=0}^{n_e}\binom{n_e}{e}(1-b)^e b^{n_e-e}\sum_{f=0}^{n_f}\binom{n_f}{f}(1-b)^f b^{n_f-f}\cdot \\
\Bigg[O\left(\exp(-2c_1)\right) + O\left(\exp(-2c_2)\right) + \\
\sum_{j:|j-\mu_1|<\sqrt{c_1 f}}\binom{f}{j}b^j(1-b)^{f-j}\sum_{k:|k-\mu_2|<\sqrt{c_2 e}}\binom{e}{k}b^k(1-b)^{e-k}\cdot \\
\left(b\Psi_2(e,f,j,k+1) + (1-b)\Psi_2(e,f,j,k)\right)\Bigg].
\end{aligned} \tag{6}$$

As j and k concentrate around their means, Ψ_2 will approach its value at those means. Let

$$\varepsilon_1(j,k,u) = \Psi_2(e,f,j,k+u) - \Psi_2(e,f,\mu_1,\mu_2+u)$$

be the difference of Ψ_2 from its value at μ_1 and $\mu_2 + u$, where $u \in \{0,1\}$ indicates if u's output is observed.

Ψ_2 is non-increasing in j and is non-decreasing in k:

$$\begin{aligned}
D_j\Psi_2 &= -\frac{(1+e)(1+f)(1+e-k)p^u_{d_{|\Delta|}}p^u_d}{\left(\begin{array}{l}(1+f)(1+e-k)p^u_{d_{|\Delta|}} + \\ (1+f-j-u)(p^u_d(e+1) + (1-p^u_d-p^u_{d_{|\Delta|}})(1+e-k))\end{array}\right)^2} \\
&\leq 0. \\
D_k\Psi_2 &= \frac{(1+e)(1+f-j)p^u_d(p^u_{d_{|\Delta|}}(1+f) + (1-p^u_d-p^u_{d_{|\Delta|}})(1+f-j))}{\left(\begin{array}{l}(1+f)(1+e-k-u)p^u_{d_{|\Delta|}} + \\ (1+f-j)((1+e)p^u_d + (1+e-k-u)(1-p^u_d-p^u_{d_{|\Delta|}}))\end{array}\right)^2} \\
&\geq 0.
\end{aligned}$$

Because the signs of these derivatives are constant, the magnitude of ε_1 is largest when j and k are as large or as small as possible. We can therefore bound the magnitude of ε_1 with

$$\begin{aligned}
&\max_{\substack{\sigma\in\{-1,1\}\\ u\in\{0,1\}}}\left(\left|\varepsilon_1\left(\mu_1+\sigma\sqrt{c_1 f},\mu_2+\sigma\sqrt{c_2 e},u\right)\right|\right) \\
&\quad= \max_{\substack{\sigma\in\{-1,1\}\\ u\in\{0,1\}}}\left|\Psi_2(e,f,\mu_1+\sigma\sqrt{c_1 f},\mu_2+\sigma\sqrt{c_2 e}+u) - \Psi_2(e,f,\mu_1,\mu_2+u)\right| \\
&\quad= O\left(\sqrt{c_1/f}\right) + O\left(\sqrt{c_2/e}\right),
\end{aligned}$$

where the second line follows from a simple expansion of Ψ_2 according to Equation 3.2. We use this estimate to approximate the value of Ψ_2:

$$\begin{aligned}\Psi_2(e,f,j,k+u) &= \Psi_2(e,f,\mu_1,\mu_2+u) + \varepsilon_1(j,k,u) \\ &= \Psi_2(e,f,\mu_1,\mu_2+u) + O\left(\sqrt{c_1/f}\right) + O\left(\sqrt{c_2/e}\right).\end{aligned}$$

We set $c_1 = \log(f)/4$ and $c_2 = \log(e)/4$, and then Equation 6 becomes

$$\begin{aligned}E[\Psi|X_D(u)=d] &= b(1-b)p_d^u + b^2 + \\ &(1-b)\sum_{e=0}^{n_e}\binom{n_e}{e}(1-b)^e b^{n_e-e}\sum_{f=0}^{n_f}\binom{n_f}{f}(1-b)^f b^{n_f-f}\cdot \\ &\Big[b\Psi_2(e,f,\mu_1,\mu_2+1) + (1-b)\Psi_2(e,f,\mu_1,\mu_2) + \\ &\quad O\left(\sqrt{\log(f)/f}\right) + O\left(\sqrt{\log(e)/e}\right)\Big].\end{aligned} \tag{7}$$

e and f in this expression are binomially distributed. Let $\mu_3 = n_e(1-b)$ be the mean of e and $\mu_4 = n_f(1-b)$ be the mean of f. By applying the Chernoff bound to the sum over e, setting the tails to start at $\min(b,1-b)n_e/2$ from μ_3, we can see that

$$\sum_{e=0}^{n_e}\binom{n_e}{e}(1-b)^e b^{n_e-e}\sum_{f=0}^{n_f}\binom{n_f}{f}(1-b)^f b^{n_f-f} O\left(\sqrt{\log(e)/e}\right) = O\left(\sqrt{\log(n_e)/n_e}\right).$$

We can similarly show that

$$\sum_{e=0}^{n_e}\binom{n_e}{e}(1-b)^e b^{n_e-e}\sum_{f=0}^{n_f}\binom{n_f}{f}(1-b)^f b^{n_f-f} O\left(\sqrt{\log(f)/f}\right) = O\left(\sqrt{\log(n_f)/n_f}\right).$$

For the remaining terms inside both sums, approximate the sums over e and f using the Chernoff bound by setting the tails to be those terms more than $\sqrt{c_3 n_e}$ from μ_3 and more than $\sqrt{c_4 n_f}$ from μ_4, respectively. This yields

$$\begin{aligned}E[\Psi|X_D(u)=d] &= b(1-b)p_d^u + b^2 + \\ &O\left((log(n_e)/n_e)^{-1/2}\right) + O\left((log(n_f)/n_f)^{-1/2}\right) + O\left(e^{-2c_3}\right) + O\left(e^{-2c_4}\right) + \\ &(1-b)\sum_{e:|e-\mu_3|<\sqrt{c_3 n_e}}\binom{n_e}{e}(1-b)^e b^{n_e-e}\sum_{f:|f-\mu_4|<\sqrt{c_4 n_f}}\binom{n_f}{f}(1-b)^f b^{n_f-f}\cdot \\ &\quad [b\Psi_2(e,f,\mu_1,\mu_2+1) + (1-b)\Psi_2(e,f,\mu_1,\mu_2)].\end{aligned} \tag{8}$$

As e and f concentrate around their means, Ψ_2 will approach its value at those means. Let

$$\varepsilon_2(e,f,u) = \Psi_2(e,f,\mu_1,\mu_2+u) - \Psi_2(\mu_3,\mu_4,\mu_1,\mu_2+u)$$

be the difference of Ψ_2 from its value at $e = \mu_3$ and $f = \mu_4$, $u \in \{0,1\}$. $\Psi_2(e, f, \mu_1, \mu_2)$ in non-decreasing with respect to e:

$$D_e\Psi_2(e,f,\mu_1,\mu_2) = \frac{(1+(1-b)f)bp_d^u((f+1)(1-p_d^u) - fb(1-p_d^u - p_{d_{|\Delta|}}^u))}{\left(\begin{array}{l}(1+(1-b)f)(1+(1-b)e)+\\(1+(1-b)f)(be)p_d^u+\\bf(1+(1-b)e+u)p_{d_{|\Delta|}}^u\end{array}\right)^2}$$
$$\geq 0.$$

$\Psi_2(e, f, \mu_1, \mu_2 + 1)$ is non-increasing with respect to e:

$$D_e\Psi_2(e,f,\mu_1,\mu_2) = \frac{(1+(1-b)f)(1-b)p_d^u(fb(1-p_{d_{|\Delta|}}^u - p_d^u) - (f+1)(1-p_d^u))}{\left(\begin{array}{l}((1-b)f)(1+(1-b)e)+\\(1+(1-b)f)(be+1)p_d^u+\\bf((1-b)e)p_{d_{|\Delta|}}^u\end{array}\right)^2}$$
$$\leq 0.$$

$\Psi_2(e, f, \mu_1, \mu_2 + u)$, $u \in \{0,1\}$, is non-increasing with respect to f:

$$D_f\Psi_2(e,f,\mu_1,\mu_2+u) = \frac{-b(1+e)(1+(1-b)e+u)p_d^u p_{d_{|\Delta|}}^u}{\left(\begin{array}{l}(1+(1-b)f)(1+(1-b)e+u)+\\(1+(1-b)f)(be+u)p_d^u+\\bf(1+(1-b)e+u)p_{d_{|\Delta|}}^u\end{array}\right)^2}$$
$$\leq 0.$$

Therefore, the magnitude of ε_2 is largest when e and f are as large or as small as possible. We can therefore estimate the magnitude of ε_2 with

$$\max_{\substack{\sigma\in\{-1,1\}\\u\in\{0,1\}}} \left(\left|\varepsilon_2\left(\mu_3+\sigma\sqrt{c_3n_e}, \mu_4+\sigma\sqrt{c_4n_f}, u\right)\right|\right).$$

If $n_e, n_f \neq 0$,

$$\begin{aligned}\varepsilon_2\left(\mu_3+\sigma\sqrt{c_3n_e}, \mu_4+\sigma\sqrt{c_4n_f}, u\right) =& \Psi_2(\mu_3+\sigma\sqrt{c_3n_e}, \mu_4+\sigma\sqrt{c_4n_f}, \mu_1, \mu_2+u) - \\ & \Psi_2(\mu_3, \mu_4, \mu_1, \mu_2+u)\\ =& O\left(\sqrt{c_3/n_e}\right) + O\left(\sqrt{c_4/n_f}\right).\end{aligned}$$

If $n_e = 0$, which occurs when $\alpha = 0$,

$$\begin{aligned}\varepsilon_2\left(0, \mu_4+\sigma\sqrt{c_4n_f}, u\right) &= \Psi_2(0, \mu_4+\sigma\sqrt{c_4n_f}, \mu_1, u) - \Psi_2(0, \mu_4, \mu_1, u)\\ &= O\left(\sqrt{c_4/n_f}\right).\end{aligned}$$

If $n_f = 0$, which occurs when $\alpha = 1$, the final term becomes

$$\begin{aligned}\varepsilon_2\left(\mu_3+\sigma\sqrt{c_3n_e}, 0, u\right) &= \Psi_2(\mu_3+\sigma\sqrt{c_3n_e}, 0, 0, \mu_2+u) - \Psi_2(\mu_3, 0, 0, \mu_2+u)\\ &= O\left(\sqrt{c_3/n_e}\right).\end{aligned}$$

These asymptotic estimates of ε_2 follow from a simple expansion of Ψ_2 according to Equation 3.2.

We use these estimates to approximate the value of Ψ_2 as e and f grow:

$$\begin{aligned}\Psi_2(e,f,\mu_1,\mu_2+u) &= \Psi_2(\mu_3,\mu_4,\mu_1,\mu_2+u)+\varepsilon_2(e,f,u)\\ &= \Psi_2(\mu_3,\mu_4,\mu_1,\mu_2+u)+O\left(\sqrt{c_3/n_e}\right)+O\left(\sqrt{c_4/n_f}\right).\end{aligned}$$

We set $c_3 = \log(n_e)/4$ and $c_4 = \log(n_f)/4$, and then Equation 8 becomes

$$\begin{aligned}E[\Psi|X_D(u)=d] = {} & b(1-b)p_d^u + b^2 + \\ & (1-b)\left[b\Psi_2(\mu_3,\mu_4,\mu_1,\mu_2+1)+(1-b)\Psi_2(\mu_3,\mu_4,\mu_1,\mu_2)\right] + \\ & O\left((log(n_e)/n_e)^{-1/2}\right)+O\left((log(n_f)/n_f)^{-1/2}\right).\end{aligned} \tag{9}$$

Finally, we must estimate $\Psi_2(\mu_3,\mu_4,\mu_1,\mu_2+u)$, $u\in\{0,1\}$. Assume that $0<\alpha<1$ and thus that $n_e=\alpha(n-1)$ and $n_f=(1-\alpha)(n-1)$ are both increasing with n. Then

$$\begin{aligned}\Psi_2(\mu_3,\mu_4,\mu_1,\mu_2+u) &= \Psi_2((1-b)n_e,(1-b)n_f,b(1-b)n_f,b(1-b)n_e+u)\\ &= \frac{p_d^u(1-b)^3n_en_f+c_1n_e+c_2n_f+c_3}{\left(\begin{array}{l}((1-b)^4+p_d^u(1-b)^3b+p_{d_{|\Delta|}}^u(1-b)^3b)n_en_f+\\ c_4n_e+c_5n_f+c_6\end{array}\right)}\\ &= \frac{p_d^u}{1-b+p_d^ub+p_{d_{|\Delta|}}^ub}+O(1/n_e)+O(1/n_f)+O(1/(n_en_f)),\end{aligned}$$

where $c_1,\ldots,c_6$ are some values constant in n_e and n_f. When $\alpha=0$, then $n_e=0$, and the estimate becomes

$$\begin{aligned}\Psi_2(\mu_3,\mu_4,\mu_1,\mu_2+u) &= \Psi_2(0,(1-b)n_f,b(1-b)n_f,u)\\ &= \frac{p_d^u(1-b)n_f+c_1}{((1-u)(1-b)+p_d^uu(1-b)+p_{d_{|\Delta|}}^u(1-u)b)n_f+c_2}\\ &= \frac{p_d^u(1-b)}{((1-u)(1-b)+p_d^uu(1-b)+p_{d_{|\Delta|}}^u(1-u)b)}+O(1/n_f),\end{aligned}$$

where c_1,c_2 are some values constant in n_f. When $\alpha=1$, then $n_f=0$, and the estimate becomes

$$\begin{aligned}\Psi_2(\mu_3,\mu_4,\mu_1,\mu_2+u) &= \Psi_2((1-b)n_e,0,0,b(1-b)n_e+u)\\ &= \frac{p_d^un_e+c_1}{((1-b)+p_d^ub)n_e+c_2}\\ &= \frac{p_d^u}{1-b+p_d^ub}+O(1/n_e),\end{aligned}$$

where c_1,c_2 are some values constant in n_e.

Inserting these estimates for $\Psi_2(\mu_3,\mu_4,\mu_1,\mu_2+u)$ into Equation 9 yields the theorem. □

It follows from this theorem that the worst case anonymity over user distributions occurs either when all users always visit $d_{|\Delta|}$ or when all users always visit d.

COROLLARY 3.7. $\lim_{n\to\inf} E[\Psi|X_D(u)=d]$ *is maximized either at* $\alpha=0$ *or at* $\alpha=1$.

PROOF. The case $\alpha = 1$ is larger in the limit than the case where $0 < \alpha < 1$, by Thm. 3.6, because

$$\frac{p_d^u}{1 - b + p_d^u b + p_{d_{|\Delta|}}^u b} \leq \frac{p_d^u}{1 - b + p_d^u b}.$$

□

The case $\alpha = 1$ is the worst case only when

$$p_{d_{|\Delta|}}^u \geq \frac{(1-b)(1-p_d^u)^2}{p_d^u(1+b) - b}.$$

This happens when $p_d^u \geq 1/2$ and $p_{d_{|\Delta|}}^u$ is near $1 - p_d^u$. That is, if the user is likely to visit d and the other users can't distinguish themselves too much, then it is worst to have them always visit d because the adversary will blame u.

However, we would expect $p_{d_{|\Delta|}}^u$ to be small because it is at most $1/|\Delta|$. In this case the worst-case limiting distribution has $\alpha = 0$, that is, it is worst when the other users always act very different from u by visiting $d_{|\Delta|}$. Then the expected assigned probability is about $b + (1-b)p_d^u$. This is equal to the lower bound on the anonymity metric when the adversary controls a fraction $\sqrt{b}$ of the network.

4. TYPICAL DISTRIBUTIONS

It is unlikely that users of onion routing will ever find themselves in the worst-case situation. The necessary distributions just do not resemble what we expect user behavior to be like in any realistic use of onion routing. Our worst-case analysis may therefore be overly pessimistic. To get some insight into the anonymity that a typical user of onion routing can expect, we consider a more realistic set of users' destination distributions in which each user selects a destination from a common Zipfian distribution. This model of user behavior is used by Shmatikov and Wang [2006] to analyze relationship anonymity in mix networks and is motivated by observations that the popularity of sites on the web follows a Zipfian distribution.

Let each user select his destination from a common Zipfian distribution p: $p_{d_i} = 1/(\mu i^s)$, where $s > 0$ and $\mu = \sum_{i=1}^{|\Delta|} 1/i^s$. It turns out that the exact form of the distribution doesn't matter as much as the fact that it is common among users.

THEOREM 4.1. *When $p^v = p^w$, for all $v, w \in U$,*

$$E[\Psi | X_D(u) = d] = b^2 + (1 - b^2)p_d^u + O(1/n)$$

PROOF. Let p be the common destination distribution. The expected assigned probability can be expressed as:

$$E[\Psi | X_D(u) = d] = b^2 + b(1-b)p_d^u + (1-b)\sum_{s=1}^{n} b^{n-s}(1-b)^{s-1}\sum_{t=0}^{s}(1-b)^{s-t}b^t\binom{n-1}{s-1}\cdot \left[\binom{s-1}{t-1}\sum_{\Delta \in D^t : \Delta_1 = d}\prod_{i=2}^{t} p_{\Delta_i}\Psi_4(s,\Delta) + \binom{s-1}{t}\sum_{\Delta \in D^t}\prod_{i=1}^{t} p_{\Delta_i}\Psi_4(s,\Delta)\right]. \quad (10)$$

Here, s represents the size of the set of users with unobserved inputs, t represents the size of the subset of those s users that also have observed outputs, Δ represents the t observed destinations, and $\Psi_4(s, \Delta)$ is the posterior probability. In this situation,

Ψ is unambiguous given s and Δ. Let $\Delta_d = |\{x \in \Delta : x = d\}|$. Ψ_4 can be expressed simply as:

$$\begin{aligned}\Psi_4(s,\Delta) &= \frac{\Delta_d(s-1)^{|\Delta|-1} + p_d(s-1)^{|\Delta|}}{s^{|\Delta|}} \\ &= (\Delta_d + p_d(s-t))/s.\end{aligned}$$

The sum

$$\sum_{\Delta \in D^t : \Delta_1 = d} \prod_{i=2}^{t} p_{\Delta_i} \Psi_4(s,\Delta)$$

in Equation 10 calculates the expectation for Ψ_4 conditioned on s and t. The expression for Ψ_4 shows that this expectation depends linearly on the expected value of Δ_d. Δ_d's expectation is simply $1 + p_d(t-1)$, because one destination in this case is always d, and each of the other $t-1$ is d with probability p_d. The sum

$$\sum_{\Delta \in D^t} \prod_{i=1}^{t} p_{\Delta_i} \Psi_4(s,\Delta)$$

in Equation 10 similarly depends linearly on the expectation of Δ_d, which in this case is $p_d t$.

With these observations, it is a straightforward calculation to show that the sum over t in Equation 10 is simply

$$b\frac{p_d(s-1)+1}{s} + (1-b)p_d.$$

We insert this into Equation 10 and simplify:

$$\begin{aligned}E[\Psi | X_D(u) = d] =& b^2 + b(1-b)p_d^u + \\ & (1-b)\sum_{s=1}^{n} b^{n-s}(1-b)^{s-1}\binom{n-1}{s-1}\left[b\frac{p_d(s-1)+1}{s} + (1-b)p_d\right] \\ =& b^2 + b(1-b)p_d^u + \\ & (1-b)\left[b\left(p_d + \frac{(1-p_d)(1-(1-b)^{n+1})}{b(n+1)}\right) + (1-b)p_d\right] \\ =& b^2 + (1-b^2)p_d^u + O(1/n).\end{aligned}$$

□

Our results show that the expected value of the anonymity metric is close to $b^2 + (1-b^2)p_d^u$ for large populations, which matches the lower bound shown in Thm. 3.3. This fact also justifies somewhat using a simple analysis that does not take into account the effect on anonymity of the behavior of the whole user population.

5. RELATED WORK

Ours is not the first formalization of anonymous communication. Early formalizations used communicating sequential processes [Schneider and Sidiropoulos 1996], graph theory and possible worlds [Hughes and Shmatikov 2004], and epistemic logic [Syverson and Stubblebine 1999; Halpern and O'Neill 2005]. These works focused primarily on formalizing the high-level concept of anonymity in communication. For this reason, they applied their formalisms to toy examples or systems that are of limited practical application and can only provide very strong forms of anonymity, *e.g.*, dining-cryptographers networks. Also, with the exception of Halpern and O'Neill [2005], they

have at most a limited ability to represent probability and probabilistic reasoning. We have focused in [Feigenbaum et al. 2007] on formalizing a widely-used, practical, low-latency system.

Halpern and O'Neill [2005] give a general formulation of anonymity in systems that applies to our model. They describe a "runs-and-systems" framework that provides semantics for logical statements about systems. They then give several logical definitions for varieties of anonymity. It is straightforward to apply this framework to the network model and protocol that we give in [Feigenbaum et al. 2007]. Our possibilistic definitions of sender anonymity, receiver anonymity, and relationship anonymity then correspond to the notion of "minimal anonymity" as defined in their paper. The other notions of anonymity they give are generally too strong and are not achieved in our model of onion routing.

Later formalizations of substantial anonymous communication systems [Camenisch and Lysyanskaya 2005; Mauw et al. 2004; Wikström 2004] have not been directly based on the design of deployed systems and have focused on provability without specific regard for applicability to an implemented or implementable design. Also, results in these papers are for message-based systems: each message is constructed to be processed as a self-contained unit by the appropriate router, typically using the generally available public encryption key for that router. Such systems typically employ mixing, changing the appearance and decoupling the ordering of input to output messages at the router to produce anonymity locally [Chaum 1981]. Onion routing, on the other hand, is circuit based: before passing any messages with user content, onion routing first lays a circuit through the routers that provides those routers the keys to be used in processing the actual messages. Mixing can be combined with onion routing in various ways [Reed et al. 1998], although this is not typical [Dingledine et al. 2004]. Such circuit creation facilitates bidirectional, low-latency coommunication and has been an identifying feature of onion routing since the first public use of the phrase [Goldschlag et al. 1996]. Thus, while illuminating and important works on anonymous communication, the formalizations above are not likely to be applicable to low-latency communications, and, despite the title of [Camenisch and Lysyanskaya 2005], are not analyses of onion routing.

Circuit construction has been done in various ways throughout the history of onion routing. In the first version of onion routing [Goldschlag et al. 1996], and other early versions [Reed et al. 1998; Goldberg and Shostack 2001], after a user selects a sequence of onion routers from a publicly-known set, the user then creates a circuit through this sequence using an *onion*, a data structure effectively composed only of layers with nothing in the middle. There is one public-key-encrypted layer for each hop in the circuit, the decryption of which contains the identity of the next hop in the circuit (if there is one) and keying material for passing data over the established circuit. In later protocols, such as used in Cebolla [Brown 2002] and Tor [Dingledine et al. 2004], the circuit is built via a telescoping protocol that extends the circuit hop-by-hop, using the existing circuit for each extension. For all of these, each hop only communicates with the routers before and after it in the sequence, and the messages are encrypted once for each router in the circuit so that no additional information leaks about the identities of the other routers or the destination of the circuit. Cryptographic techniques are used so that message forgery is countered. Some later designs returned to the non-interactive circuit construction of the original [Øverlier and Syverson 2007; Kate et al. 2007]. It is trivial to see that all of these fit directly within our model.

Some versions of onion routing, such as those that do iterative discovery of onion routers via a DHT [Freedman and Morris 2002; Mittal and Borisov 2009; McLachlan et al. 2009], will not fit within our model without some extensions that we do not pursue herein. This is because the probability of first-last router choice and router com-

promise within a circuit can no longer be assumed to be independent. Some anonymity protocols that do not use onion routing may nonetheless also fit within our model, appropriately extended. For example, in Crowds [Reiter and Rubin 1998], the adversary can learn from observing the first and last routers, but the connection to the first router does not automatically identify the source. On the other hand the destination is always known to every router in the circuit. The probability that an observed circuit predecessor is the source can thus be combined with the observed destination and the a priori source-destination probability distribution.

In this paper, we add probabilistic analysis to the framework of [Feigenbaum et al. 2007]. Other works have presented probabilistic analysis of anonymous communication [Reiter and Rubin 1998; Shmatikov 2004; Wright et al. 2004; Danezis 2003; Danezis and Serjantov 2004; Mathewson and Dingledine 2004; Kesdogan et al. 1998] and even of onion routing [Syverson et al. 2000]. The work of Shmatikov and Wang [2006] is particularly similar to ours. It calculates relationship anonymity in mix networks and incorporates user distributions for selecting destinations. However, with the exception of [Shmatikov 2004], these have not been formal analyses. Also, whether for high-latency systems such as mix networks, or low-latency systems, such as Crowds and onion routing, many of the attacks in these papers are some form of intersection attack. In an intersection attack, one watches repeated communication events for patterns of senders and receivers over time. Unless all senders are on and sending all the time (in a way not selectively blockable by an adversary) and/or all receivers are receiving all the time, if different senders have different receiving partners, there will be patterns that arise and eventually differentiate the communication partners. It has long been recognized that no system design is secure against a long-term intersection attack. Several of these papers set out frameworks for making that more precise. In particular, [Danezis 2003], [Danezis and Serjantov 2004], and [Mathewson and Dingledine 2004] constitute a progression towards quantifying how long it takes (in practice) to reveal traffic patterns in realistic settings.

We are not concerned herein with intersection attacks. We are effectively assuming that the intersection attack is done. The adversary already has a correct distribution of a user's communication partners. We are investigating the anonymity of a communication in which a user communicates with one of those partners in the distribution. This follows the anonymity analyses performed in much of the literature [Kesdogan et al. 1998; Mauw et al. 2004; Reiter and Rubin 1998; Syverson et al. 2000], which focus on finding the source and destination of an individual communication. Our analysis differs in that we take into account the probabilistic nature of the users' behavior. Probabilistic anonymity metrics used previously include, when applied to our situation, the probability assigned to the correct destination [Reiter and Rubin 1998], the entropy of the destination distribution [Díaz et al. 2002; Serjantov and Danezis 2002], and maximum probability within the destination distribution [Tóth et al. 2004], where the distribution in each case is a conditional distribution given the adversary's view.

We expect this to have potential practical applications. For example, designs for shared security-alert repositories to facilitate both forensic analysis for improved security design and quicker responses to widescale attacks have been proposed [Lincoln et al. 2004]. A participant in a shared security-alert repository might expect to be known to communicate with it on a regular basis. Assuming reports of intrusions, etc., are adequately sanitized, the concern of the participant should be to hide when it is that updates from that participant arrive at the repository, *i.e.*, which updates are likely to be from that participant as opposed to others.

6. CONCLUSIONS AND FUTURE WORK

We expect each user of an anonymity network to have a pattern of use. In order to make guarantees to the user about his anonymity, we need to take this into account when modeling and analyzing the system, especially in light of previous research that indicates that an adversary can learn these usage patterns given enough time.

We perform such an analysis on onion routing. Onion routing is a successful design used, in the form of the Tor system, by hundreds of thousands of people to protect their security and privacy. But, because it was designed to be practical and because theory in this area is still relatively young, the formal analysis of its privacy properties has been limited.

We perform our analysis using a simple black-box model in the UC framework. We justify this model by showing that it information-theoretically provides the same anonymity as the onion routing protocol formalized by Feigenbaum et al. [2007] and by recognizing that it can be UC-realized. Furthermore, it should lend itself to the analysis of other anonymity protocols expressed within the UC framework. We investigate the relationship anonymity of users and their destinations in this model and measure it using the probability that the adversary assigns to the correct destination of a given user after observing the network.

Our anonymity analysis first shows that a simple, standard approximation to the expected value of the anonymity metric provides a lower bound on it. Then we consider the worst-case set of user behaviors to give an upper bound on the expected value. We show that, in the limit as the number of users grows, a user's anonymity is worst either when all other users choose destinations he is unlikely to visit, because that user becomes unique and identifiable, or when that user chooses a destination that all other users prefer, because the adversary mistakes the group's choices for the user's choice. This worst-case anonymity with an adversary that controls a fraction b of the routers is comparable to the best-case anonymity against an adversary that controls a fraction $\sqrt{b}$.

The worst case is unlikely to be the case for any users; so we investigate anonymity under a more reasonable model of user behavior suggested in the literature. In it, users select destinations from a common Zipfian distribution. Our results show that, in this case and in any case with a common distribution, the expected anonymity tends to the best possible, *i.e.* the adversary doesn't usually gain that much knowledge from the other users' actions.

Our anonymity analysis provides some justification for the non-rigorous analysis that is typically used with onion-routing security. However, it also shows that, in the worst case, user behaviors can interact to degrade anonymity to a surprising degree; therefore, in unusual situations this factor should be taken into account.

Future work includes extending this analysis to other types of anonymity (such as sender anonymity), extending it to other anonymity networks, and learning more about the belief distribution of the adversary than just its mean. A big piece of the attack we describe is in learning the users' destination distribution, about which only a small amount of research, usually on simple models, has been done. The speed with which an adversary can perform this stage of the attack is crucial in determining the validity of our attack model and results.

In response to analyses such as that of Øverlier and Syverson [2006], the current Tor design includes entry guards by default for all circuits. Roughly, this means that, since about January 2006, each Tor client selects its first onion router from a small set of nodes that it randomly selects at initialization. The rationale is that communication patterns of individuals are what need to be protected. If an entry guard is compromised, then the percentage of compromised circuits from that user is much higher.

But, without entry guards, it appears that whom that user communicates with and even at what rate can be fairly quickly learned by an adversary owning a modest percentage of the Tor nodes anyway. If no entry guard is compromised, then no circuits from that user will ever be linked to him. However, if a user expects to be targeted by a network adversary that can control nodes, he can expect his entry guards ultimately to be attacked and possibly compromised. If the destinations he chooses that are most sensitive are rarely contacted, he may thus be better off choosing first nodes at random. How can we know which is better? Extending our analysis to include entry guards will allow us to answer or at least further illuminate this question.

Our model also assumes that client connections to the network are such that the initial onion router in a circuit can tell that it is initial for that circuit. This is true for the overwhelming majority of traffic on the Tor network today, because most users run clients that are not also onion routers. However, for circuits that are initiated at a node that runs an onion router, a first node cannot easily tell whether it is the first node or the second—without resorting to other attacks of unknown efficacy, *e.g.*, monitoring latency of traffic moving in each direction in response to traffic moving in the other direction. Thus, that initiating edge of the black box is essentially fuzzy. Indeed, this was originally the only intended configuration of onion routing for this reason [Goldschlag et al. 1996]. The addition of clients that do not also function as routers was a later innovation that was added to increase usability and flexibility [Reed et al. 1998; Syverson et al. 2000]. Similarly, peer-to-peer designs such as Crowds [Reiter and Rubin 1998] and Tarzan [Freedman and Morris 2002] derive their security even more strongly from the inability of the first node to know whether it is first or not. Thus, extending our model and analysis to this case will make it still more broadly applicable.

A. APPENDIX

Let $\tilde{f}$ be as defined in Lemma 3.5.

LEMMA A.1. $D^2_{s_{d_i}} \tilde{f} \geq 0$.

PROOF. Let $i = s_{d_i}$ and $\mu = \nu - m$ for simplicity. Then

$$\tilde{f} = \frac{(\nu + i\mu)(\nu + (\nu - i)\mu)}{p^u_{d_j}\nu(\nu + i\mu)(1 - i + \nu) + (1 + i)p^u_{d_i}\nu(\nu - i\mu + \nu\mu) + \beta(\nu + i\mu)(\nu - i\mu + \nu\mu)}.$$

The second derivative of $\tilde{f}$ can be expressed as

$$D^2_{s_{d_i}} \tilde{f} = \frac{N}{D},$$

where

$$\begin{aligned} N = -\Bigg(& (2(i+j)(-i-j+\mu) \\ & \left(-i^3(p^u_{d_i} - p^u_{d_j})\mu^3((i+j)(p^u_{d_i} + p^u_{d_j}) + \beta\mu) + \right. \\ & 3i^2(i+j)\mu^2(p^u_{d_i} + p^u_{d_j} + p^u_{d_i}\mu)((i+j)(p^u_{d_i} + p^u_{d_j}) + \beta\mu) - \\ & 3i(i+j)^2\mu((i+j)(p^u_{d_i} + p^u_{d_j}) + \beta\mu)\left(-p^u_{d_j} + p^u_{d_i}(1+\mu)^2\right) + \\ & (i+j)^3\left((i+j)(p^u_{d_i})^2(1+\mu)^3 + p^u_{d_j}((i+j)p^u_{d_j} + \beta\mu) + \right. \\ & \left.\left.\left. p^u_{d_i}\left(\beta\mu(1+\mu)^3 + p^u_{d_j}(2+\mu)\left(-i-j+2\mu+(1+i+j)\mu^2\right)\right)\right)\right)\Bigg) \end{aligned}$$

and

$$D = \Bigg((i+j)^2(p^u_{d_i} + ip^u_{d_i} + p^u_{d_j} + jp^u_{d_j} + \beta) + \\ (i+j)(i(p^u_{d_j} + \beta) + j(p^u_{d_i} + ip^u_{d_i} + ip^u_{d_j} + \beta))\mu + ij\beta\mu^2\Bigg)^3,$$

substituting $(i+j)$ for ν. D is clearly positive. Therefore we must just show that N is non-negative.

We collect terms in N by the coefficients p_{d_i}, p_{d_j}, and β:

$$\begin{aligned}
&2p_{d_j}\beta(i+j)(i+j-\mu)\mu(i+j+i\mu)^3+\\
&2p_{d_i}\beta(i+j)(i+j-\mu)\mu(i+j+j\mu)^3+\\
&2(p^u_{d_j})^2(i+j)^2(i+j-\mu)(i+j+i\mu)^3+\\
&2(p^u_{d_i})^2(i+j)^2(i+j-\mu)(i+j+j\mu)^3+\\
&2p^u_{d_i}p^u_{d_j}(i+j)^2(i+j-\mu)(i+j)(2+\mu)\cdot\\
&\qquad\left(i^2\left(-1+\mu^2\right)+j\left(\mu(2+\mu)+j\left(-1+\mu^2\right)\right)+i\left(\mu(2+\mu)-j\left(2+\mu^2\right)\right)\right).
\end{aligned}$$

The coefficients of the terms in p_{d_i} and p_{d_j} are clearly positive because $i+j=\nu\geq\nu-m=\mu$.

If we collect the remaining terms by i and j, we get

$$\begin{aligned}
&i^3\left((p^u_{d_i})^2+(p^u_{d_j})^2(1+\mu)^3+p^u_{d_i}p^u_{d_j}\left(-2-\mu+2\mu^2+\mu^3\right)\right)+\\
&j^3\left((p^u_{d_j})^2+(p^u_{d_i})^2(1+\mu)^3+p^u_{d_i}p^u_{d_j}\left(-2-\mu+2\mu^2+\mu^3\right)\right)+\\
&i^2p^u_{d_i}p^u_{d_j}\mu(2+\mu)^2+\\
&j^2p^u_{d_i}p^u_{d_j}\mu(2+\mu)^2+\\
&2ijp^u_{d_i}p^u_{d_j}\mu(2+\mu)^2+\\
&3i^2j\left((p^u_{d_i})^2(1+\mu)+(p^u_{d_j})^2(1+\mu)^2-p^u_{d_i}p^u_{d_j}(2+\mu)\right)+\\
&3ij^2\left((p^u_{d_j})^2(1+\mu)+(p^u_{d_i})^2(1+\mu)^2-p^u_{d_i}p^u_{d_j}(2+\mu)\right).
\end{aligned}$$

The coefficients for the i^3 and j^3 terms are clearly non-negative when $\mu\geq 1$. When $\mu=0$, observe that the coefficients become $(p^u_{d_i}-p^u_{d_j})^2\geq 0$. The coefficients for the i^2, j^2, and ij terms are also clearly non-negative.

To show that the i^2j term is non-negative, we use the fact that p_{d_i} and p_{d_j} are probabilities that sum to at most one. Let $p_{d_j}=\zeta-p_{d_i}$, $0\leq\zeta\leq 1$. Then the coefficient of i^2j becomes a quadratic function of p_{d_i} with positive second derivative. Its minimum is at

$$p_{d_i}=\frac{4\zeta+5\zeta\mu+2\zeta\mu^2}{2(2+\mu)^2}.$$

The coefficient evaluated at this point is

$$\frac{\zeta^2\mu\left(8+11\mu+4\mu^2\right)}{4(2+\mu)^2},$$

which is non-negative. Therefore, the whole i^2j term is non-negative.

Similarly, for the ij^2 term, we look at its coefficient as a function of p_{d_i} with $p_{d_j} = \zeta - p_{d_i}$. It is also a quadratic function with positive second derivative. Its minimum is found at

$$\frac{4\zeta + 3\zeta\mu}{2(2+\mu)^2}.$$

The coefficient evaluated at this point is

$$\frac{\zeta^2\mu(8+\mu(11+4\mu))}{4(2+\mu)^2},$$

which is non-negative. Therefore, the whole ij^2 term is non-negative. This implies that N is non-negative, and thus that $D^2_{s_{d_i}}\tilde{f}$ is non-negative. □

REFERENCES

BACKES, M., GOLDBERG, I., KATE, A., AND MOHAMMADI, E. 2012. Provably secure and practical onion routing. In *Proceedings of the 25th IEEE Computer Security Foundations Symposium (CSF 2012)*. Forthcoming.

BAUER, K., MCCOY, D., GRUNWALD, D., KOHNO, T., AND SICKER, D. 2007. Low-resource routing attacks against Tor. In *Proceedings of the Workshop on Privacy in the Electronic Society (WPES 2007)*. 11–20.

BEIMEL, A. AND DOLEV, S. 2003. Buses for anonymous message delivery. *Journal of Cryptology 16,* 1, 25–39.

BROWN, Z. 2002. Cebolla: Pragmatic IP anonymity. In *Proceedings of the 2002 Ottawa Linux Symposium*.

CAMENISCH, J. AND LYSYANSKAYA, A. 2005. A formal treatment of onion routing. In *Proceedings of CRYPTO 2005*. 169–187.

CANETTI, R. 2000. Universally composable security: A new paradigm for cryptographic protocols. Cryptology ePrint Archive, Report 2000/067. http://eprint.iacr.org/.

CHAUM, D. 1981. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM 4,* 2, 84–88.

CHAUM, D. 1988. The dining cryptographers problem: Unconditional sender and recipient untraceability. *Journal of Cryptology: The Journal of the International Association for Cryptologic Research 1,* 1, 65–75.

CORRIGAN-GIBBS, H. AND FORD, B. 2010. Dissent: accountable anonymous group messaging. In *Proceedings of the 17th ACM conference on Computer and communications security (CCS 2010)*. 340–350.

DANEZIS, G. 2003. Statistical disclosure attacks: Traffic confirmation in open environments. In *Proceedings of Security and Privacy in the Age of Uncertainty (SEC 2003)*. 421–426.

DANEZIS, G. AND SERJANTOV, A. 2004. Statistical disclosure or intersection attacks on anonymity systems. In *Proceedings of 6th Information Hiding Workshop (IH 2004)*. 293–308.

DÍAZ, C., SEYS, S., CLAESSENS, J., AND PRENEEL, B. 2002. Towards measuring anonymity. In *Proceedings of the 2nd Privacy Enhancing Technologies Workshop (PET 2002)*. 54–68.

DINGLEDINE, R., MATHEWSON, N., AND SYVERSON, P. 2004. Tor: The second-generation onion router. In *Proceedings of the 13th USENIX Security Symposium*. 303–319.

FEIGENBAUM, J., JOHNSON, A., AND SYVERSON, P. 2007. A model of onion routing with provable anonymity. In *Proceedings of the 11th Financial Cryptography and Data Security Conference (FC 2007)*. 57–71.

FREEDMAN, M. J. AND MORRIS, R. 2002. Tarzan: A peer-to-peer anonymizing network layer. In *Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS 2002)*. 193–206.

GOLDBERG, I. AND SHOSTACK, A. 1999. Freedom 1.0 security issues and analysis. White paper, Zero Knowledge Systems, Inc. November.

GOLDBERG, I. AND SHOSTACK, A. 2001. Freedom network 1.0 architecture and protocols. White paper, Zero Knowledge Systems, Inc. October. The attributed date is that printed at the head of the paper. The cited work is, however, superceded by documents that came before Oct. 2001. The appendix indicates a change history with changes last made November 29, 1999. Also, in [Goldberg and Shostack 1999] the same authors refer to a paper with a similar title as an "April 1999 whitepaper".

GOLDSCHLAG, D. M., REED, M. G., AND SYVERSON, P. F. 1996. Hiding routing information. In *Information Hiding: First International Workshop*. 137–150.

HALPERN, J. Y. AND O'NEILL, K. R. 2005. Anonymity and information hiding in multiagent systems. *Journal of Computer Security 13,* 3, 483–514.

HERRMANN, D., WENDOLSKY, R., AND FEDERRATH, H. 2009. Website fingerprinting: attacking popular privacy enhancing technologies with the multinomial naïve-bayes classifier. In *Proceedings of the 2009 ACM workshop on Cloud Computing Security (CCSW '09)*. 31–42.

HOPPER, N., VASSERMAN, E. Y., AND CHAN-TIN, E. 2010. How much anonymity does network latency leak? *ACM Transactions on Information and System Security 13,* 2, 1–28.

HUGHES, D. AND SHMATIKOV, V. 2004. Information hiding, anonymity and privacy: A modular approach. *Journal of Computer Security 12,* 1, 3–36.

KATE, A., ZAVERUCHA, G., AND GOLDBERG, I. 2007. Pairing-based onion routing. In *Privacy Enhancing Technologies: 7th International Symposium, (PET 2007)*. 95–112.

KESDOGAN, D., AGRAWAL, D., AND PENZ, S. 2002. Limits of anonymity in open environments. In *Proceedings of the 5th Information Hiding Workshop (IH 2002)*. 53–69.

KESDOGAN, D., EGNER, J., AND BÜSCHKES, R. 1998. Stop-and-go MIXes: Providing probabilistic anonymity in an open system. In *Proceedings of the 2nd Information Hiding Workshop (IH 1998)*. 83–98.

LINCOLN, P., PORRAS, P., AND SHMATIKOV, V. 2004. Privacy-preserving sharing and correlation of security alerts. In *Proceedings of the 13th USENIX Security Symposium*. 239–254.

LOESING ET AL. 2011. Tor metrics portal. https://metrics.torproject.org/.

LYNCH, N. A. 1996. *Distributed Algorithms*. Morgan Kaufmann Publishers Inc.

MATHEWSON, N. AND DINGLEDINE, R. 2004. Practical traffic analysis: Extending and resisting statistical disclosure. In *Proceedings of the 4th Privacy Enhancing Technologies workshop (PET 2004)*. 17–34.

MAUW, S., VERSCHUREN, J., AND DE VINK, E. 2004. A formalization of anonymity and onion routing. In *Proceedings of the 9th European Symposium on Research in Computer Security (ESORICS 2004)*. 109–124.

MCLACHLAN, J., TRAN, A., AND HOPPER, N. 2009. Scalable onion routing with Torsk. In *Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS'09)*. ACM Press, 590–599.

MITTAL, P. AND BORISOV, N. 2009. ShadowWalker: Peer-to-peer anonymous communication using redundant structured topologies. In *Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS'09)*. 161–172.

MURDOCH, S. J. 2006. Hot or not: Revealing hidden services by their clock skew. In *Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS'06)*. 27–36.

MURDOCH, S. J. AND DANEZIS, G. 2005. Low-cost traffic analysis of Tor. In *Proceedings of the 2005 IEEE Symposium on Security and Privacy (S& P 2005)*. 183–195.

NAMBIAR, A. AND WRIGHT, M. 2006. Salsa: A structured approach to large-scale anonymity. In *Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS 2006)*.

ØVERLIER, L. AND SYVERSON, P. 2006. Locating hidden servers. In *Proceedings of 2006 IEEE Symposium on Security and Privacy (S& P 2006)*. 100–114.

ØVERLIER, L. AND SYVERSON, P. 2007. Improving efficiency and simplicty of Tor circuit establishment and hidden services. In *Privacy Enhancing Technologies: 7th International Symposium (PET 2007)*. 134–152.

PFITZMANN, A. AND HANSEN, M. 2000. Anonymity, unobservability, and pseudonymity: A consolidated proposal for terminology. Draft.

REED, M. G., SYVERSON, P. F., AND GOLDSCHLAG, D. M. 1998. Anonymous connections and onion routing. *IEEE Journal on Selected Areas in Communications 16,* 4, 482–494.

REITER, M. AND RUBIN, A. 1998. Crowds: Anonymity for web transactions. *ACM Transactions on Information and System Security 1,* 1, 66–92.

SCHNEIDER, S. AND SIDIROPOULOS, A. 1996. CSP and anonymity. In *Proceedings of the 1st European Symposium on Research in Computer Security (ESORICS 1996)*. 198–218.

SERJANTOV, A. AND DANEZIS, G. 2002. Towards an information theoretic metric for anonymity. In *Proceedings of the 2nd Privacy Enhancing Technologies Workshop (PET 2002)*. 41–53.

SHMATIKOV, V. 2004. Probabilistic model checking of an anonymity system. *Journal of Computer Security 12,* 3-4, 355–377.

SHMATIKOV, V. AND WANG, M.-H. 2006. Measuring relationship anonymity in mix networks. In *Proceedings of the 5th ACM Workshop on Privacy in the Electronic Society (WPES 2006)*. 59–62.

SYVERSON, P., REED, M., AND GOLDSCHLAG, D. 2000. Onion routing access configurations. In *Proceedings of the DARPA Information Survivability Conference and Exposition (DISCEX 2000)*. 34–40.

SYVERSON, P., TSUDIK, G., REED, M., AND LANDWEHR, C. 2000. Towards an Analysis of Onion Routing Security. In *Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*. 96–114.

SYVERSON, P. F. AND STUBBLEBINE, S. G. 1999. Group principals and the formalization of anonymity. In *Proceedings of the 1st World Congress on Formal Methods (FM'99), Vol. I*. 814–833.

TÓTH, G., HORNÁK, Z., AND VAJDA, F. 2004. Measuring anonymity revisited. In *Proceedings of the 9th Nordic Workshop on Secure IT Systems*. 85–90.

WIKSTRÖM, D. 2004. A universally composable mix-net. In *First Theory of Cryptography Conference (TCC 2004)*. 317–335.

WRIGHT, M. K., ADLER, M., LEVINE, B. N., AND SHIELDS, C. 2004. The predecessor attack: An analysis of a threat to anonymous communications systems. *ACM Transactions on Information and System Security 7,* 4, 489–522.