AnoA: A Differential Anonymity Framework

Michael Backes, Aniket Kate, Praveen Manoharan, Sebastian Meiser, Esfandiar Mohammadi

CISPA, Saarland University

***Abstract*—Protecting individuals' privacy in online communications has become a challenge of paramount importance. To this end, anonymous communication (AC) protocols such as the widely used Tor network have been designed to provide anonymity to their participating users. In this work we present AnoA: a generic framework for defining, analyzing, and quantifying anonymity properties for AC protocols. AnoA relies on a novel relaxation of the notion of (computational) differential privacy, and thereby enables a unified quantitative analysis of well-established anonymity properties, such as sender anonymity and recipient anonymity.**

Moreover we present MATor, a tool for rigorously assessing the degree of anonymity in the Tor network. MATor utilizes AnoA and outputs guarantees for the Tor network that explicitly address how user anonymity is impacted by real-life characteristics of actually deployed Tor, such as its path selection algorithm, Tor consensus data, and the preferences and the connections of the user.

***Keywords*-anonymity analysis; differential privacy; Tor; provable privacy**

I. Introduction

Protecting individuals' privacy in online communications has become a challenge of paramount importance. A wide variety of privacy enhancing technologies, comprising many different approaches, have been proposed to solve this problem. Privacy enhancing technologies, such as anonymous communication (AC) protocols, seek to protect users' privacy by anonymizing their communication over the Internet. Employing AC protocols has become increasingly popular over the last decade. This popularity is exemplified by the success of the Tor network [28].

There has been a substantial amount of previous work [26], [10], [23], [24], [19], [17], [25], [9], [13], [14], [16], [1], [15] on analyzing the anonymity provided by various AC protocols such as dining cryptographers network (DC-net) [8], Crowds [22], mix network (Mixnet) [7], and onion routing (e.g., Tor) [21]. However, most of the previous works only consider a single anonymity property for a particular AC protocol under a specific adversary scenario. Previous frameworks such as [18] only guarantee anonymity for a symbolic abstraction of the AC, not for its cryptographic realization. Moreover, while some existing works like [15] consider an adversary with access to *a priori* probabilities for the behavior of users, there is still no work that is capable of dealing with an adversary that has arbitrary auxiliary information about user behavior.

A. Contributions

In this work we present the novel anonymity analysis framework AnoA. In AnoA we define and analyze anonymity properties of AC protocols. Our anonymity definition is based on a novel generalization of differential privacy, a notion for privacy preserving computation that has been introduced by Dwork et al. [11], [12]. The strength of differential privacy resides in a strong adversary that has maximal control over two adjacent settings that it has to distinguish. However, applying differential privacy to AC protocols seems impossible. While differential privacy does not allow for leakage of (potentially private) data, AC protocols inherently leak to the recipient the data that a sender sends to this recipient. We overcome this contradiction by generalizing the adjacency of settings between which an adversary has to distinguish. We introduce an explicit *anonymity function* (corresponding to a notion of adjacency) α that characterizes whether two settings are considered adjacent or not. In contrast to previous work on anonymity properties, our formulation of anonymity properties enables the adversary to choose the messages as long as the adjacent challenge inputs carry the same messages. Moreover, AnoA is compatible with simulation-based composability frameworks, such as UC [6]. In particular, for all protocols that are securely abstracted by an ideal functionality [2], our definitions allow an analysis of these protocols in a purely information theoretical manner.

We formalize the well-established notions of sender anonymity, recipient anonymity, and relationship anonymity in our framework, by introducing appropriate anonymity functions.

As an analysis technique, we present MATor: the first system to derive sender, recipient and relationship anonymity guarantees based on Tor's real-life characteristics, such as its actual path selection strategy. Our anonymity definitions are modular and take into account actual Tor consensus data and user preferences (e.g., concerning the path selection algorithm).

We apply our analysis technique to recent Tor Metrics data [27] to perform a comprehensive analysis of Tor's

Adaptive AnoA **Challenger** $\text{Ch}(\mathcal{P}, \alpha, n, b)$

Upon message$(\mathsf{input}, r^* = (U, S, m, \mathsf{sid}))$

1: RunProtocol$(r^*, 0)$

Upon message $(\mathsf{challenge}, r_0, r_1, \Psi)$

1: **if** $\Psi \notin \{1, \ldots, n\}$ **then** abort
2: **else if** $\Psi \in \mathbb{T}$ **then**
3: Retrieve $st := st_\Psi$
4: **if** $st = \mathsf{over}$ **then** abort
5: **else** $st := \mathsf{fresh}$ and add Ψ to $\mathbb{T}$
6: Compute $(r^*, st_\Psi) \leftarrow \alpha(st, r_0, r_1, b)$
7: Store st_Ψ.
8: RunProtocol(r^*, Ψ)

RunProtocol$(r = (U, S, m, \mathsf{sid}), \Psi)$

1: **if** $\neg\exists y$ such that $(\mathsf{sid}, y, \Psi) \in S$ **then**
2: Let $\mathsf{sid}_{\text{real}} \leftarrow \{0,1\}^k$; Store $(\mathsf{sid}, \mathsf{sid}_{\text{real}}, \Psi)$ in S.
3: **else** $\mathsf{sid}_{\text{real}} := y$
4: Run $\mathcal{P}$ on $r = (U, S, m, \mathsf{sid}_{\text{real}})$ and forward all messages that are sent by $\mathcal{P}$ to the adversary $\mathcal{A}$ and send all messages by the adversary to $\mathcal{P}$.

Figure 1. Adaptive AnoA Challenger

anonymity guarantees. To this end, we conduct a large scale evaluation of different path selection algorithms for a broad variety of trust models, ranging from simple adversaries that compromise a given number of Tor nodes, over geographic adversaries (e.g., adversaries that compromise all nodes within certain countries), up to complex adversary models that follow economic reasoning. Due to space restrictions we focus on Tor's standard path selection algorithm and on two adversaries: one that compromises a number of k arbitrary nodes and a simple geographical adversary that compromises all nodes within a certain country.

II. The AnoA Framework

AnoA provides a parametric definition for classes of anonymity definitions in the spirit of adaptive challenge-response games. The results of these games are used to define quantitative anonymity guarantees, intuitively by bounding the probability for every efficient (probabilistic, polynomial-time) adversary to distinguish two scenarios s_0 and s_1 that reflect the considered anonymity notion. For instance, sender anonymity is captured by the (in-)ability of an adversary to distinguish two different possible senders of one message that is sent to the same destination.

Technically, all anonymity notions in AnoA are defined in terms of a so-called *adaptive* AnoA *challenger* Ch that interacts with an arbitrary probabilistic polynomial-time Turing machines $\mathcal{A}$, called the *adversary*. This challenger is not only parametric in the protocol under consideration, but also in the anonymity notion, i.e., there exists a unique challenger that can be instantiated to various scenarios. We show the formal definition of this challenger in Figure 1. The challenger receives as input a description of the protocol $\mathcal{P}$, the anonymity notion under consideration expressed as a function α, the number of allowed challenges n, and a challenge bit b. The goal of the adversary $\mathcal{A}$ will be to distinguish between interactions with $\text{Ch}(\mathcal{P}, \alpha, n, 0)$ and interactions with $\text{Ch}(\mathcal{P}, \alpha, n, 1)$. In the following, we write Ch_α^b instead of $\text{Ch}(\mathcal{P}, \alpha, n, b)$ if $\mathcal{P}$ and n are clear from the context. The challenger Ch_α^b accepts two different kinds of inputs. First, it allows the adversary to control actions of users within the network, i.e., the adversary can trigger users to send specific messages to specific recipients. This is modeled by an *input* message of the form $(\mathsf{input}, r = (U, S, m, \mathsf{sid}))$, where U and S denote sender and recipient of the message m, respectively, and where sid is a session identifier that enables communication across individual invocations. Upon receiving this input, the challenger executes the underlying protocol to perform this send request, abbreviated by a function call to RunProtocol. Second, the challenger accepts up to n challenge messages of the form $(\mathsf{challenge}, r_0, r_1, \Psi)$. For every such challenge, uniquely identified by the tag Ψ, the challenger maintains a state st_Ψ that is fresh (for newly started challenges), over (for completed challenges), or contains information about the (ongoing) challenge. For every challenge that is not in the state over, the challenger calls the anonymity function α on the state st_Ψ, the messages r_0 and r_1 and the challenge bit b. Depending on which anonymity notion the function α describes, it computes an action r^* that a user should execute within the network. The challenger finally executes the protocol $\mathcal{P}$ for user action r^*, while additionally making sure that all protocol invocations from input messages and challenges messages can be told apart. Moreover, the challenger allows protocol participants to send messages to the challenger and vice versa.

After defining the challenger, corresponding probability bounds (ϵ, δ) of a given adversary $\mathcal{A}$ are defined as follows.

Definition 1 ((n, ε, δ)-α-IND-CDP)**.** *Let Ch be the challenger from Figure 1. The protocol $\mathcal{P}$ is (n, ε, δ)-α-IND-CDP against an adversary class A, where $\varepsilon \geq 0$ and $0 \leq \delta \leq 1$, if for all* PPT*-adversaries $\mathcal{A} \in A$:*

$$\Pr\left[0 \leftarrow \langle \mathcal{A} || \text{Ch}(\mathcal{P}, \alpha, n, b) \rangle\right] \leq e^\varepsilon \Pr\left[0 \leftarrow \langle \mathcal{A} || \text{Ch}(\mathcal{P}, \alpha, n, 1-b) \rangle\right] + \delta.$$

It thus provides a quantitative assessment of anonymity in the sense of computational differential privacy [20], and it thus allows differentiating between distinguishing events and small comparative gains of information.

A. Anonymity Definitions

Sender anonymity. Sender anonymity intuitively captures the anonymity of a user U against a (possibly malicious) destination. By definition of the scenario, the destination S sees the exit node and tries to find out which one of two possible users U_1 and U_2 it is interacting with.

Recall that the different anonymity notions are defined as parameters α in the overall challenge-response game. Sender anonymity α_{SSA} is formally defined as $\alpha_{\text{SSA}}(st, r_0 = (\mathcal{S}_0, \mathcal{R}_0, m_0, \_), r_1 = (\mathcal{S}_1, \_, \_, \_), b) :=$

if $st = \mathsf{fresh} \vee st = (\mathcal{S}_0, \mathcal{S}_1)$ **then**
output $((\mathcal{S}_b, \mathcal{R}_0, m_0, 1), st := (\mathcal{S}_0, \mathcal{S}_1))$

Recipient anonymity. Recipient anonymity intuitively captures the anonymity of a user U against a (possibly malicious) ISP of this user. By definition of the scenario, the ISP sees both the user U who starts a connection and the entry node of every circuit. Its goal is to identify which of two possible destinations S_1 and S_2 the user communicates with. Recipient anonymity α_{SRA} is defined as $\alpha_{\text{SRA}}(st, r_0 = (\mathcal{S}_0, \mathcal{R}_0, m_0, \_), r_1 = (\_, \mathcal{R}_1, m_1, \_), b) :=$

if $(st = \mathsf{fresh} \vee st = \mathcal{S}_0) \wedge |m_0| = |m_1|$ **then**
output $((\mathcal{S}_0, \mathcal{R}_b, m_b, 1), st := \mathcal{S}_0)$

Realistic Adversaries. We model adversaries that resemble realistic scenarios, by introducing the concept of adversary classes. These adversary classes allow for restricting the strong AnoA adversary to the scenario of interest.

Technically, an adversary class $A(\cdot)$ is a wrapper that restricts the adversary $\mathcal{A}$ in its possible output behavior, and thus, in its knowledge about the world. Technically, it is a PPT machine $A(\mathcal{A})$ that internally runs the adversary $\mathcal{A}$ and forwards all messages that are sent from a compromised node to the adversary $\mathcal{A}$ and vice versa. We refer to the full version [3] for a detailed description of these technical adversary classes.

Moreover, instead of only considering k-of-n adversaries (adversaries that freely compromise k arbitrary nodes within a set of n nodes), we aim to capture more sophisticated adversary classes for different types of adversarial corruptions, such as corruption based on geo-locality, bandwidth, or cost-functions for every node n. Defining appropriate classes within the underlying framework then ensures that the adversary compromises nodes according to the considered restrictions.

Instead of defining an individual class for each of these considered adversary scenarios, we define a parametric adversary class that we call *budget-adversary* class, out of which we will instantiate all relevant individual adversary classes. The budget-adversary is parametric in a given *cost function* f that assigns costs to every node n within the Tor network, and in a *budget* G that the adversary may spend to corrupt nodes.

Definition 2 (Budget-Adversary)**.** *A* budget-adversary class $A_f^B(\cdot)$*, or* budget-adversary *for short, for a given cost function f and budget B is a Turing machine that upon input of an adversary machine $\mathcal{A}$ behaves as follows:*

- $A_f^B(\mathcal{A})$ *keeps track of a budget b and initializes it as $b := B$.*
- $A_f^B(\mathcal{A})$ *internally simulates $\mathcal{A}$ and forwards all messages from $\mathcal{A}$ to the protocol and the challenger and vice versa, with the exception of compromise requests, as specified in the next bullet.*
- *Whenever $\mathcal{A}$ sends a command* **compromise**(x) *for a node x, $A_f^B(\mathcal{A})$ verifies that $f(x) \leq b$. If this holds, it forwards the command to the protocol and reduces the budget $b := b - f(x)$. Otherwise it responds with an error message.*

III. MATor: Measuring anonymity guarantees

We developed the anonymity measurement tool MATor [4] which computes the impact of the path selection algorithm on the anonymity of a user. The tools uses the actual Tor metrics data for the measurement and enables the specification of a wide variety of adversary classes. Using our theoretical framework AnoA, we prove that the results of MATor are secure.

Because of space constraints we present only two (very simple) example adversaries, namely an adversary that compromises a fixed number of k nodes and a geographical adversary that compromises all nodes within a country. In a technical report [5], we conducted extensive experiments with more complex adversary classes such as bandwidth-compromising adversaries, botnet-adversaries and adversaries that have a monetary budget. These experiments are founded in the AnoA framework and supported by MATor as well.

A. k-of-n adversaries

We begin with the k-of-N adversary model in which the adversary may compromise up to k nodes of its choice. This worst-case adversary is useful for estimating the maximal impact that a collaboration of a certain number of participants can have on the anonymity within the Tor network. Such an adversary typically compromises the nodes with the largest weight and thus we expect this adversary to be stronger whenever the trust is not distributed evenly over the nodes. Formally, we instantiate our budget adversary class to model that the adversary may compromise k arbitrary nodes (out of all $N = |\mathcal{N}|$ Tor nodes), independent of their properties, by using $f^{k\text{-of-}N}(x) := 1$ for all nodes $x \in \mathcal{N}$. The adversary class is then $A^k_{f^{k\text{-of-}N}}$.

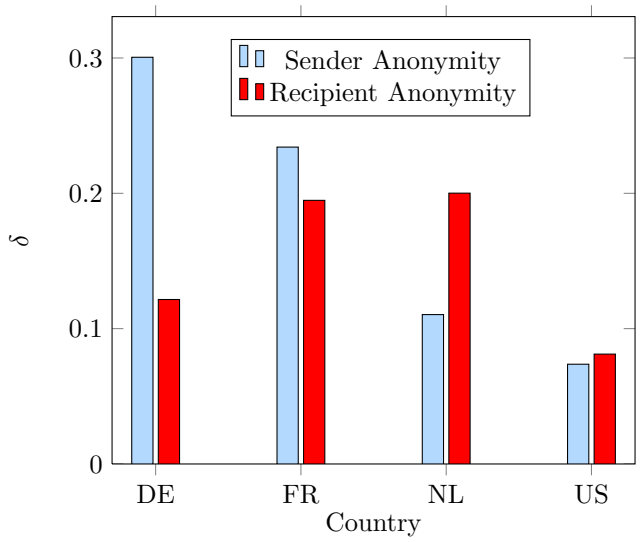

Figure 2. Advantage δ of a country based adversary, depending on the country for which it compromises all nodes, for the countries France (FR), Germany (DE), Netherlands (NL) and United States (US), ordered by the guarantee for Tor's path selection algorithm – results for sender anonymity and recipient anonymity. [5]

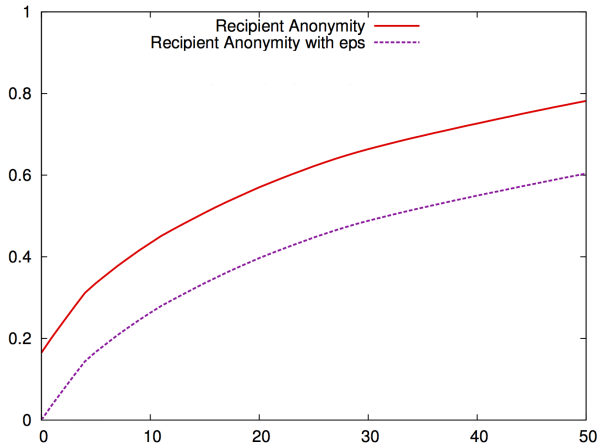

Figure 3. Impact of a multiplicative factor on recipient anonymity [4]. We chose values for ε of $\varepsilon = 0.25$ for one setting and $\varepsilon = 0$ for the other setting. The graph depicts the value for δ computed by the monitor for both path selection algorithms. The scenarios are: HTTPS + IRC vs. HTTPS on 5 February 2014.

B. Geographic adversaries

We define a geographical adversary that is completely independent of bandwidth or a specific budget, as it can compromise all nodes that are located within a specific country. Such an adversary model can reflect the fear of a user that an oppressive regime tries to deanonymize the communication to find out either the sender or the recipient of the communication. In such scenarios, the user might fear that all nodes that lie within the geographical (or jurisdictional) border of a country can be compromised (e.g., forced to reveal information) by the regime. We formalize this intuition of geographical adversaries by first introducing a slight variant of budget adversary classes $A^{B=1}_{f^\Pi}$ for boolean predicates Π, where f^Π is defined as:

$$f^\Pi(x) = \begin{cases} 0 & \text{if } \Pi(x) = 1 \\ \infty & \text{otherwise} \end{cases}$$

We then instantiate the predicate Π by country predicates Π_C for countries C, defined as

$$\Pi_\mathrm{C}(x) := \begin{cases} 1 & \text{if } x.\mathsf{country} = \mathrm{C} \\ 0 & \text{otherwise} \end{cases}$$

By choosing a country C, we can formally define an adversary that can eavesdrop on all nodes within this country, e.g., the adversary $A^{B=1}_{f^{\Pi_\mathrm{NL}}}$ with

$$f^{\Pi_\mathrm{NL}}(x) = \begin{cases} 0 & \text{if } x.\mathsf{country} = \mathrm{NL} \\ \infty & \text{otherwise} \end{cases}$$

compromises all nodes within the Netherlands. In our analysis, we instantiate Π_C for all countries C we wish to analyze.

These geographical cost functions assume that the adversary can (basically for free) compromise all nodes within the country, but it cannot compromise other nodes. Such an adversary allows to evaluate how much impact a country has on the Tor network in terms of anonymity. We show the advantage δ of the geographical adversary for the four countries Germany, France, Netherlands and US in Figure 2. (This selection was made since, to improve readability, we have ordered the countries by the advantage of PSTor and selected the top four countries.)

C. The impact of a multiplicative factor

The definition of AnoA introduces a multiplicative factor in addition to the normal additive factor (that often suffices to describe the success probability of an adversary). This factor allows for accounting for various events in which an adversary might gain information that may even lead to a non-negligible advantage without overestimating these events.

The experiments (see Figure 3) show that such a factor often only plays a minor role, as the probability to completely deanonymize a user is for most settings higher than the probability to just learn some information about them. Recipient anonymity in a setting with a weaker adversary, that compromises no, or only a very limited amount of nodes presents a noteworthy exception. Recall that for recipient anonymity we assume that the ISP of the user is compromised, which means that the adversary can see which entry node the user connects to. For different ports the probability of choosing these

entry nodes, however, will be different, because they might also be possible exit nodes, or related to possible exit nodes. An adversary that compromises no (only a very limited number of) nodes can have already a non-negligible advantage in guessing which port a user might choose, which can either be expressed by a multiplicative factor and a δ of zero, or by a non-negligible δ.

References

[1] E. Andrés, Miguel, Catuscia Palamidessi, Ana Sokolova, and Peter Van Rossum. Information Hiding in Probabilistic Concurrent System. *Journal of Theoretical Computer Science (TCS)*, 412(28):3072–3089, 2011.

[2] Michael Backes, Ian Goldberg, Aniket Kate, and Esfandiar Mohammadi. Provably Secure and Practical Onion Routing. In *Proc. 26st IEEE Symposium on Computer Security Foundations (CSF)*, pages 369–385, 2012.

[3] Michael Backes, Aniket Kate, Praveen Manoharan, Sebastian Meiser, and Esfandiar Mohammadi. AnoA: A Framework For Analyzing Anonymous Communication Protocols — Unified Definitions and Analyses of Anonymity Properties. available at http://www.infsec.cs.uni-saarland.de/~meiser/paper/anoa.html.

[4] Michael Backes, Aniket Kate, Sebastian Meiser, and Esfandiar Mohammadi. (Nothing else) MATor(s): Monitoring the Anonymity of Tor's Path Selection. In *Proceedings of the 21st ACM Conference on Computer and Communications Security (CCS)*, pages 513–524. ACM, 2014.

[5] Michael Backes, Sebastian Meiser, and Marcin Slowik. Your Choice MATor(s): Assessing Tor Anonymity Guarantees for Different Path Selection Algorithms and Trust Models. available
anoa3.pdf.

[6] R. Canetti. Universally Composable Security: A New Paradigm for Cryptographic Protocols. In *Proc. 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 136–145, 2001.

[7] D. Chaum. Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. *Communications of the ACM*, 4(2):84–88, 1981.

[8] David Chaum. The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability. *J. Cryptology*, 1(1):65–75, 1988.

[9] Claudia Díaz. Anonymity Metrics Revisited. In *Anonymous Communication and its Applications*, 2006.

[10] Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards Measuring Anonymity. In *Proc. 2nd Workshop on Privacy Enhancing Technologies (PET)*, pages 54–68, 2002.

[11] Cynthia Dwork. Differential Privacy. In *ICALP (2)*, pages 1–12, 2006.

[12] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating Noise to Sensitivity in Private Data Analysis. In *Proc. 10th Theory of Cryptography Conference (TCC)*, pages 265–284, 2006.

[13] J. Feigenbaum, A. Johnson, and P. F. Syverson. A Model of Onion Routing with Provable Anonymity. In *Proc. 11th Conference on Financial Cryptography and Data Security (FC)*, pages 57–71, 2007.

[14] J. Feigenbaum, A. Johnson, and P. F. Syverson. Probabilistic Analysis of Onion Routing in a Black-Box Model. In *Proc. 6th ACM Workshop on Privacy in the Electronic Society (WPES)*, pages 1–10, 2007.

[15] Joan Feigenbaum, Aaron Johnson, and Paul F. Syverson. Probabilistic Analysis of Onion Routing in a Black-Box Model. *ACM Transactions on Information and System Security (TISSEC)*, 15(3):14, 2012.

[16] Benedikt Gierlichs, Carmela Troncoso, Claudia Díaz, Bart Preneel, and Ingrid Verbauwhede. Revisiting a Combinatorial Approach toward Measuring Anonymity. In *Proc. 7th ACM Workshop on Privacy in the Electronic Society (WPES)*, pages 111–116, 2008.

[17] Joseph Y. Halpern and Kevin R. O'Neill. Anonymity and Information Hiding in Multiagent Systems. *Journal of Computer Security*, 13(3):483–512, 2005.

[18] Dominic Hughes and Vitaly Shmatikov. Information Hiding, Anonymity and Privacy: a Modular Approach. *Journal of Computer Security*, 12(1):3–36, 2004.

[19] S. Mauw, J. Verschuren, and E. de Vink. A Formalization of Anonymity and Onion Routing. In *Proc. 9th European Symposium on Research in Computer Security (ESORICS)*, pages 109–124, 2004.

[20] Ilya Mironov, Omkant Pandey, Omer Reingold, and Salil P. Vadhan. Computational Differential Privacy. In *Advances in Cryptology — CRYPTO*, volume 5677, pages 126–142, 2009.

[21] M. Reed, P. Syverson, and D. Goldschlag. Anonymous Connections and Onion Routing. *IEEE J-SAC*, 16(4):482–494, 1998.

[22] Michael K. Reiter and Aviel D. Rubin. Crowds: Anonymity for Web Transactions. *ACM Transactions on Information and System Security (TISSEC)*, 1(1):66–92, 1998.

[23] Andrei Serjantov and George Danezis. Towards an Information Theoretic Metric for Anonymity. In *Proc. 2nd Workshop on Privacy Enhancing Technologies (PET)*, pages 41–53, 2002.

[24] Vitaly Shmatikov. Probabilistic Analysis of an Anonymity System. *Journal of Computer Security*, 12(3-4):355–377, 2004.

[25] Vitaly Shmatikov and Ming-Hsiu Wang. Measuring Relationship Anonymity in Mix Networks. In *Proc. 7th ACM Workshop on Privacy in the Electronic Society (WPES)*, pages 59–62, 2006.

[26] P. Syverson, G. Tsudik, M. Reed, and C. Landwehr. Towards an Analysis of Onion Routing Security. In *Proc. Workshop on Design Issues in Anonymity and Unobservability (WDIAU)*, pages 96–114, 2000.

[27] Tor Metrics Portal. https://metrics.torproject.org/. Accessed Feb 2014.

[28] The Tor Project. https://www.torproject.org/, 2003. Accessed Feb 2014.