DE GRUYTER OPEN
Proceedings on Privacy Enhancing Technologies ; 2016 (2):40–60

Michael Backes, Sebastian Meiser*, and Marcin Slowik

Your Choice MATor(s)

Large-scale Quantitative Anonymity Assessment of Tor Path Selection Algorithms Against Structural Attacks

Abstract: In this paper, we present a rigorous methodology for quantifying the anonymity provided by Tor against a variety of structural attacks, i.e., adversaries that corrupt Tor nodes and thereby perform eavesdropping attacks to deanonymize Tor users. First, we provide an algorithmic approach for computing the *anonymity impact* of such structural attacks against Tor. The algorithm is parametric in the considered path selection algorithm and is, hence, capable of reasoning about variants of Tor and alternative path selection algorithms as well. Second, we present formalizations of various instantiations of structural attacks against Tor and show that the computed anonymity impact of each of these adversaries indeed constitutes a worst-case anonymity bound for the cryptographic realization of Tor. Third, we use our methodology to conduct a *rigorous, large-scale evaluation* of Tor's anonymity which establishes worst-case anonymity bounds against various structural attacks for Tor and for alternative path selection algorithms such as DistribuTor, SelekTOR, and LASTor. This yields the first rigorous anonymity comparison between different path selection algorithms. As part of our analysis, we quantify the anonymity impact of a path selection transition phase, i.e., a small number of users decides to run an alternative algorithm while the vast majority still uses the original one. The source code of our implementation is publicly available.

Keywords: Tor, anonymity quantification, Tor's path selection, anonymous communication, rigorous guarantees

DOI 10.1515/popets-2016-0004
Received 2015-08-31; revised 2015-12-02; accepted 2015-12-02.

Michael Backes: CISPA, Saarland University & MPI-SWS, E-mail: backes@cs.uni-saarland.de
***Corresponding Author: Sebastian Meiser:** CISPA, Saarland University, E-mail: meiser@cs.uni-saarland.de
Marcin Slowik: CISPA, Saarland University, E-mail: me@marandil.pl

1 Introduction

The Internet has grown from a small network to an omnipresent backbone of our society that manages and enables commercial, social, and political activities worldwide. The indisputable benefits of this transformation are, however, accompanied by novel privacy threats: User activities are constantly tracked and profiled, and the collected information is used for targeted advertising by industry and for dragnet surveillance at the planetary scale by almost omnipotent governmental agencies. In fact, new revelations about governmental observations and large-scale user profiling by various companies make it into the news with distressing regularity.

As a result, public interest in anonymous communication systems has vastly increased, and millions of users have started to use anonymizing proxies and VPNs to anonymously browse the web. In particular the Tor network [5, 15] has received tremendous attention in this respect, both as an end-user solution, currently serving more than 1.5 million people from all over the world, and as a building block for further anonymizing systems such as the privacy-preserving operating system Tails [3]. In Tor, a user connects to a sequence of three proxies (out of a set of currently more than six thousand volunteer proxies, called *nodes*), and thereby forms a so-called *Tor circuit.* The anonymity provided by this construction inherently depends on a user's trust in these nodes and on the likelihood of selecting trusted or compromised nodes in the circuit generation phase.

Assessing this degree of anonymity for different trust assumptions has spawned a multitude of research on analyzing the impact that any compromised Tor node can have on the anonymity of a user. However, most existing works provide no rigorous bounds on the provided anonymity. They are instead restricted to empirical analyses and simulations that strive to measure the anonymity impact of malicious Tor nodes; or they only consider coarse-grained, all-or-nothing attacks that would result in immediate deanonymization. The few recent approaches that aim at rigorously quantifying the anonymity of Tor against compromised nodes are restricted in scope in that they only consider simplistic adversaries and in that they are specific to individual variants of Tor's still evolving node selection algorithm. There is a lack of generic, comprehensive framework that allows for assessing anonymity against a wide selection of structural attacks (i.e., corrupting Tor nodes and thereby performing eavesdropping attacks) and for comparing these variants with each other as well as with recently proposed, alternative path selection algorithms such as DistribuTor [10], SelekTOR [22], and LASTor [7].

1.1 Our Contribution

In this paper, we present a rigorous methodology for quantifying the anonymity impact of compromised Tor nodes for any variant of Tor's path selection algorithm and alternative path selection algorithms. Our contribution is twofold: we present an algorithmic approach for computing the worst-case anonymity impact of adversaries that compromise Tor nodes, and we evaluate the anonymity impact of such adversaries for different path selection algorithms.

Computing the Anonymity Impact. Algorithmically quantifying the anonymity impact of adversaries that compromise Tor nodes in a sound manner constitutes a challenging task. We strive to go beyond the prevalently considered all-or-nothing anonymity assessments, which only consider attacks in which the adversary immediately observes both ends of a communication and, hence, achieves an immediate deanonymization. A more careful investigation shows that additional conclusions that reduce anonymity can be drawn when corrupting *any* node, and these conclusions are no less influential. Tor's node selection strategies can depend on properties of the sender (e.g., by using a specific algorithm) and of the recipient (imposing requirements, such as the supported ports of the connection) of a communication. Hence, we first carefully model which *observations* any subset of Tor nodes can make. After that, we show how to compute the *anonymity impact* of such observations for the commonly considered three anonymity notions: sender anonymity, recipient anonymity, and relationship anonymity. We model arbitrary structural adversaries, in the sense of corrupting nodes in order to mount eavesdropping attacks, using the novel concept of *budget adversaries*. Budget adversaries have a certain budget B and a cost function f that assigns a cost to every Tor node. They can compromise an arbitrary subset of Tor nodes as long as the aggregated node cost does not exceed the budget. We show that budget adversaries can be instantiated in various ways to model different structural attacks against Tor, ranging from k-collusion adversaries that corrupt a certain number of nodes to adversaries that corrupt nodes based on geographic locations and adversaries that corrupt nodes subject to monetary constraints. Next, we show how to compute the worst-case anonymity impact of a budget adversary based on the anonymity impact of the observations of all individual nodes. We then prove that this computed anonymity impact for every budget adversary indeed constitutes a worst-case anonymity bound for an idealized version of Tor in the AnoA framework [8] – a recent framework for proving quantitative bounds for anonymous communication protocols; moreover, these bounds are tight for adversaries that observe exactly one Tor node. Finally, we show that our bounds also hold for the cryptographic realization of Tor, up to a negligible factor.

Large-scale Evaluation of Tor's Anonymity. We demonstrate the applicability of our methodology to large-scale analyses by performing the to date largest *rigorous anonymity evaluation* of the Tor network. Based on recent Tor Metrics data [5], we compute anonymity bounds for Tor's standard path selection algorithm and several variants thereof, including LASTor [7], SelekTOR [22], DistribuTor [10], and the uniform routing strategy against a broad variety of structural adversaries. These include including k-collusion adversaries (compromising a certain number of Tor nodes), bandwidth-adversaries (compromising Tor nodes of a certain total bandwidth), predicate adversaries (compromising all nodes based on a predicate check, such as their geographic location or the Tor version they are running), and monetary adversaries that pertain to economic considerations (compromising selected nodes based on a given price function). Our evaluation yields the first rigorous, quantitative anonymity comparison between different path selection algorithms. Moreover, we explicitly cover the impact of a path selection transition phase, i.e., a small number of users already uses an alternative path selection algorithm, while the majority still relies on Tor's standard path selection algorithm. Our evaluation shows that such pioneers are highly vulnerable, even against adversaries that compromise only few Tor nodes. Moreover, we explicitly evaluate the advantage of adversaries that mount so-called guard detection attacks. We consider this to be of particular interest since Tor recently implemented a novel guard-selecting strategy [14] that restricts users to a single guard over a 9-month period. Our results in particular show that for a set of guard nodes accounting for approximately 10% of the entry bandwidth, an adversary that solely inspects the recipient without compromising any nodes can already distinguish between pairs of senders with a 2.12% advantage, in extreme cases up to 7.65%. The source code of our implementation is available [2].

1.2 Related Work

The Tor literature is rich on proposals for new path selection algorithms. Some propose to increase the anonymity of the users [7, 10, 16] by reducing the attack vectors of an adversary that controls part of the Internet infrastructure or part of the Tor network. Others propose to increase the performance (i.e., expected latency and throughput) of the Tor network [26]. Several existing works analyze or measure the anonymity of users within the Tor network. We categorize these into works that strive for rigorous worst-case guarantees and works that empirically determine anonymity.

In the category of rigorous worst-case guarantees, [8, 17, 18] analyze Tor based on an idealized functionality and probabilistic methods. All these works assume that the path selection algorithm chooses nodes uniformly at random (which Tor does not) and, hence, do not provide rigorous guarantees. Moreover, these formalizations ignore subtle, yet potentially influential, differences in adversarial observations whenever different senders or recipients impact the probabilities of the selected Tor circuits, e.g., because they use specific parameters for Tor or even alternative path selection algorithms. Closely related to this paper, Backes et al. [10] formally analyze Tor's path selection algorithm and provide an anonymity monitor, which takes into account real-life parameters such as the number of Tor nodes and their entrusted weight within the Tor consensus. Their formalization, however, is limited to Tor's path selection algorithm and DistribuTor (their own closely related alternative path selection algorithm) and to simplistic k-collusion adversaries. Moreover, the anonymity guarantees they provide significantly overestimate the adversary's impact on Tor, as they use imprecise heuristics for calculating the anonymity impact of malicious Tor nodes on the overall guarantee. In contrast, we precisely characterize the anonymity impact of observations and only slightly (and explicitly) over-approximate our worst-case guarantees for budget-adversaries in order to improve the performance of the computation. Furthermore, our methodology for calculating guarantees directly applies to all variants of Tor's path selection algorithm and to all alternative path selections.

In the category of empirical analyses without rigorous anonymity guarantees, Johnson et al. [20] present a simulation of the Tor network, based on a probabilistic (bandwidth-based) adversary that compromises a certain percentage of Tor's bandwidth. Murdoch and Watson [21] present an analysis of proposed path selection algorithms against (bandwidth-based) adversaries that can inject malicious nodes into the Tor network, subject to a specific adversarial budget. Their work inspired the formalization of our budget adversary, with the difference that our adversary compromises existing nodes instead of adding new nodes. Other works strive to analyze Tor against network-level adversaries, which we consider a highly interesting, yet orthogonal, problem. In this area, Jaggard et al. [19] propose a path selection adaptation based on network trust to reduce the impact of network adversaries. Wacek et al. [25] analyze the impact of path selection algorithms on anonymity and performance by simulating a significant fraction of the Tor network, and then they analyze the anonymity of various path selection algorithms against (AS-level) network adversaries. The amount of analyses based on simulations and measurements further underlines the importance of a rigorous approach for quantitatively assessing the anonymity of Tor's path selection algorithm and comparing it against alternative variants.

2 Observations and their Anonymity Impact

In this section, we show how to compute bounds on the anonymity provided by Tor in the presence of an adversary that can observe the communication at certain Tor nodes and potentially, at the sender or the recipient of a communication. To begin with, we characterize the possible circuit observations that such an adversary can make in Tor, introduce the anonymity notions considered in this paper, and show how to quantitatively assess the impact of circuit observations for each of these notions (Section 2.1). We then define several adversary classes that reflect different structural corruptions (Section 2.2) and show how to compute anonymity bounds for Tor against such adversaries (Section 2.3). We later instantiate these adversary classes with conceivable real-life adversaries, thereby obtaining concrete anonymity bounds for various adversarial settings.

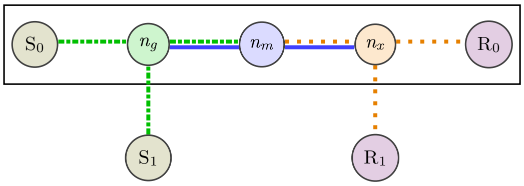


Fig. 1. Upper part: A Tor circuit consisting of a sender S_0, a guard node n_g, a middle node n_m, an exit node n_x, and a recipient R_0. Colored lines for each node depict the observations this node can make. Lower part: Further observations for an alternative sender S_1 and for an alternative recipient R_1 that will be used to define different anonymity notions.

In this section, we concentrate on the algorithmic aspects of providing anonymity bounds and use the terms *anonymity notion* and *adversary's advantage*, which are key for the overall anonymity assessment, only in an informal way. We give formal semantics to these terms in Section 3 and then rigorously prove the correctness of our computed bounds based on this semantics.

2.1 Defining Observations and their Anonymity Impact

Observing any nodes involved in a Tor circuit enables the adversary to draw certain conclusions about the sender and/or the recipient of the circuit, thereby reducing their degree of anonymity.

Some of these conclusions are straightforward and result in immediate deanonymization: if the guard node can be observed, the sender is trivially deanonymized as the origin of the communication; similarly, observing the exit node unveils the recipient as the destination of the communication. Existing papers are typically limited to such all-or-nothing observations. However, additional conclusions can be drawn when corrupting *any* node, and these conclusions are no less influential. For instance, if the adversary observes or knows that the sender communicates over a specific port, all exit nodes can be excluded that do not support this port choice, and hence, any communication that involves excluded exit nodes cannot originate from that sender. Moreover, excluding exit nodes influences the probability of which nodes are being selected as guard or middle nodes in this circuit by Tor's path selection algorithm. (The selection takes so-called family relationships and further constraints into account.) Technically, this means that the a-priori probability distribution over circuits induced by Tor's path selection algorithm is now replaced by an a-posteriori distribution that is conditioned on the observations of the adversary. This enables the adversary to draw further conclusions and to thereby reduce anonymity.

Observations. Now we define which observations an adversary is able to make if certain nodes are considered corrupted. We consider a distinguished symbol, denoted $\perp$, that reflects that an observation at a certain position in the Tor circuit cannot be made. Further, we define the overall impact on anonymity if a given set of nodes is considered under adversarial control.

Definition 1 (Observations and Circuits)**.** *For two senders* S_0*,* S_1*, two recipients* R_0*,* R_1*, and the set of all Tor nodes* $\mathcal{N}$*, we define the set of* circuits *between these senders and recipients as* $\mathcal{C}_{S_0,S_1,R_0,R_1,\mathcal{N}} := \{S_0, S_1\} \times \mathcal{N}^3 \times \{R_0, R_1\}$ *and the set of* observations *as* $Obs_{S_0,S_1,R_0,R_1,\mathcal{N}} := \{S_0, S_1, \perp\} \times (\mathcal{N} \cup \{\perp\})^3 \times \{R_0, R_1, \perp\}$ *for the distinguished symbol* $\perp$*. We omit the subscripts if they are clear from the context; hence, write* $\mathcal{C}$ *and* Obs*.*

For a set $N \subseteq \mathcal{N} \cup \{S_0, R_0\}$, we now define the observations $\mathcal{O}[N](c) \in Obs$ made by N within a considered Tor circuit c. Intuitively, whenever a node $n \in N$ is part of the circuit c, then this node as well as its successor and predecessor can be identified, see Figure 1. If $n \in \{S_0, R_0\}$, then sender and guard node (if $n = S_0$) or exit node and recipient (if $n = R_0$) can be identified.

Definition 2 (Circuit Observations)**.** *For two senders* S_0, S_1*, two recipients* R_0, R_1*, and for* $N \subseteq \mathcal{N} \cup \{S_0, R_0\}$*, the* circuit observation *of* N *is a function* $\mathcal{O}[N] : \mathcal{C} \to Obs$ *and is defined as follows. For* $c = (S_a, n_g, n_m, n_x, R_b) \in \mathcal{C}$*, we have* $\mathcal{O}[N](c) := (n_1, \ldots n_5)$ *with*

- $n_1 := S_a$ *if* $\{S_0, n_g\} \cap N \neq \emptyset$*; otherwise* $n_1 := \perp$*.*
- $n_2 := n_g$ *if* $\{S_0, n_g, n_m\} \cap N \neq \emptyset$*; otherwise* $n_2 := \perp$*.*
- $n_3 := n_m$ *if* $\{n_g, n_m, n_x\} \cap N \neq \emptyset$*; otherwise* $n_3 := \perp$*.*
- $n_4 := n_x$ *if* $\{n_m, n_x, R_0\} \cap N \neq \emptyset$*; otherwise* $n_4 := \perp$*.*
- $n_5 := R_b$ *if* $\{n_x, R_0\} \cap N \neq \emptyset$*; otherwise* $n_5 := \perp$*.*

We call N *the* observation points *of* $\mathcal{O}$*.*

Anonymity Notions. We consider three common notions of communication anonymity α in this paper: *sender anonymity* ($\alpha = \alpha_{SA}$, i.e., determine who is sending a message), *recipient anonymity* ($\alpha = \alpha_{RA}$, i.e., determine to whom a message is being sent), and *relationship anonymity* ($\alpha = \alpha_{REL}$, i.e., determine a correlation between sender and recipient). Each of these notions is defined as the (in-)ability of an adversary to distinguish two *scenarios* that differ in their involved senders and recipients. This follows the established concept of indistinguishability-based definitions in cryptography (e.g., IND-CCA secure encryption): one of these two scenarios is selected at random, a Tor circuit is created for this scenario, and the adversary is then allowed to make observations for this circuit depending on the set of corrupted nodes. The adversary knows the set-up of both scenarios, makes its observations and then has to decide which scenario it currently observes. The reduction of anonymity is then defined as the *adversary's advantage*, i.e., as the probability of correctly distinguishing both scenarios.

Each of these three notions requires its own two scenarios to define the adversary's advantage with respect to this notion. This is illustrated in Figure 1: for sender anonymity, an additional sender S_1 is considered, i.e., the two scenarios differ in the sender, but share the same recipient R_0. In addition to its observations from corrupted nodes, the adversary is allowed to observe the recipient R_0 and should be able to distinguish if the communication originates at S_0 or at S_1. Similarly, an additional recipient R_1 is considered for recipient anonymity, i.e., the two scenarios differ in the recipient, but share the same sender S_0; the adversary additionally observes the sender S_0 and tries to tell R_0 and R_1 apart. Capturing the absence of correlations to define relationship anonymity is more involved. We consider both an additional sender S_1 and an additional recipient R_1: The first relationship anonymity scenario considers the two cases that S_0 communicates with R_0 and that S_1 communicates with R_1; the second scenario considers the communication from S_0 to R_1 and from S_1 to R_0. After the scenario has been selected, one of the two described cases for this scenario is chosen uniformly at random, then a Tor circuit is created for this case and the adversary can make its observations for this circuit.

Anonymity Impact of Observations. Any circuit observation contributes information that helps an adversary to distinguish the two scenarios of the considered anonymity notion. For formally defining this *observation impact*, let $\mathsf{ps}(\mathrm{S}_a, \mathrm{R}_b)$ denote the probability distribution over Tor circuits $\mathcal{C}_{\mathrm{S}_0,\mathrm{S}_1,\mathrm{R}_0,\mathrm{R}_1,\mathcal{N}}$ induced by Tor's path selection algorithm ps (or the alternative algorithm that we consider) if $\mathrm{S}_a \in \{\mathrm{S}_0, \mathrm{S}_1\}$ creates a circuit to communicate with $\mathrm{R}_b \in \{\mathrm{R}_0, \mathrm{R}_1\}$. Then, the observation impact $\mathbf{Impact}_X^{\mathsf{obs}}(N)$ for anonymity notion α_X, the considered senders $\mathrm{S}_0, \mathrm{S}_1$ and recipients $\mathrm{R}_0, \mathrm{R}_1$, and a set of observation points $N \subseteq \mathcal{N} \cup \{\mathrm{S}_0, \mathrm{R}_0\}$, is defined as the aggregated difference of all circuit observation probabilities for the respective scenarios of the considered anonymity notion.[1]

Definition 3 (Observation Impact). *Let* $\mathrm{S}_0, \mathrm{S}_1$ *denote two senders, let* $\mathrm{R}_0, \mathrm{R}_1$ *denote two recipients, let* $N \subseteq \mathcal{N} \cup \{\mathrm{S}_0, \mathrm{R}_0\}$ *denote a set of observation points, and let* α_X *for* $X \in \{\mathrm{SA}, \mathrm{RA}, \mathrm{REL}\}$ *be an anonymity notion. Define* $\phi(Y, Z)$ *as* $Y - Z$ *if* $Y > Z$ *and* 0 *otherwise. Then,* $\mathbf{Impact}_X^{\mathsf{obs}}(N)$*, as defined in Figure 2, denotes the* observation impact *of* N *for* α_X *and* $\mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1$.

For singletons $N = \{n\}$, we write $\mathbf{Impact}_X^{\mathsf{obs}}(n)$ instead of $\mathbf{Impact}_X^{\mathsf{obs}}(\{n\})$.

2.2 Defining Structural Corruptions

In this section, we define different classes of structural adversaries that statically compromise a certain subset of Tor nodes. For conveniently reasoning about different such adversaries in a unified manner, we define the concept of a *budget adversary.*

Definition 4 (Budget Adversary). *Given a* cost function $f : \mathcal{N} \to \mathbb{N} \cup \{\infty\}$ *and a* budget $B \in \mathbb{N}$*, an adversary is called a* budget adversary $\mathcal{A}_f^B$ *if it can compromise arbitrary sets of Tor nodes* $N \subseteq \mathcal{N}$*, as long as* $\sum_{n \in N} f(n) \leq B$.

We provide several instantiations for budget adversaries.

Definition 5 (k-collusion Adversary). *A* k-collusion adversary *is a budget adversary* $\mathcal{A}_{f_{\mathsf{KofN}}}^k$ *that compromises up to k nodes of its choice, i.e.,* $f_{\mathsf{KofN}}(n) := 1$ *for* $n \in \mathcal{N}$.

Definition 6 (Predicate Adversary). *A* predicate adversary *is a budget adversary* $\mathcal{A}_{f_P}^1$ *that compromises all nodes that fulfill a given predicate* P*, i.e.,* $f_P(n) := 0$ *if* $P(n) = \mathsf{true}$*, and* $f_P(n) := \infty$ *otherwise.*

Examples of predicate adversaries include *geographic adversaries* that compromise all nodes within a certain country or a collaboration of countries, *Tor-Version adversaries* that can exploit vulnerabilities of specific versions of the Tor software and can compromise all nodes

[1] For the sake of readability, we did not explicitly include $\mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1$ and ps as additional parameters of $\mathbf{Impact}_X^{\mathsf{obs}}(N)$ but consider them clear from the context, similarly in the upcoming definitions.

$$\begin{aligned}
\mathbf{Impact}^{\mathsf{obs}}_{\mathrm{SA}}(N) &:= \sum_{o \in Obs} \phi\Big(\Pr\left[o = \mathcal{O}[N](c), c \leftarrow \mathsf{ps}(\mathrm{S}_0, \mathrm{R}_0)\right], \Pr\left[o = \mathcal{O}[N](c), c \leftarrow \mathsf{ps}(\mathrm{S}_1, \mathrm{R}_0)\right] \Big);\\
\mathbf{Impact}^{\mathsf{obs}}_{\mathrm{RA}}(N) &:= \sum_{o \in Obs} \phi\Big(\Pr\left[o = \mathcal{O}[N](c), c \leftarrow \mathsf{ps}(\mathrm{S}_0, \mathrm{R}_0)\right], \Pr\left[o = \mathcal{O}[N](c), c \leftarrow \mathsf{ps}(\mathrm{S}_0, \mathrm{R}_1)\right] \Big);\\
\mathbf{Impact}^{\mathsf{obs}}_{\mathrm{REL}}(N) &:= \sum_{o \in Obs} \phi\Big(\big(\Pr\left[o = \mathcal{O}[N](c), c \leftarrow \mathsf{ps}(\mathrm{S}_0, \mathrm{R}_0)\right] + \Pr\left[o = \mathcal{O}[N](c), c \leftarrow \mathsf{ps}(\mathrm{S}_1, \mathrm{R}_1)\right] \big)/2,\\
&\qquad\qquad \big(\Pr\left[o = \mathcal{O}[N](c), c \leftarrow \mathsf{ps}(\mathrm{S}_0, \mathrm{R}_1)\right] + \Pr\left[o = \mathcal{O}[N](c), c \leftarrow \mathsf{ps}(\mathrm{S}_1, \mathrm{R}_0)\right] \big)/2 \Big).
\end{aligned}$$

Fig. 2. Definition of $\mathbf{Impact}^{\mathsf{obs}}_{X}(N)$ to define observation impact (Definition 3)

that run this version, and *subnet adversaries* that compromise all Tor nodes within a specific IP-subnet.

Definition 7 (Bandwidth Adversary). *A resource-constrained bandwidth adversary, or* bandwidth adversary *for short, is a budget adversary $\mathcal{A}^{B}_{f_{\texttt{BW}}}$ that compromises an arbitrary set of Tor nodes with at most an overall bandwidth of B, i.e., for $n \in \mathcal{N}$, we have $f_{\texttt{BW}}(n) := n.\texttt{BW}$ for $n \in \mathcal{N}$, where $n.\texttt{BW}$ denotes the bandwidth of node n.*

This adversary model allows us to provide anonymity bounds in the presence of adversaries that manage to observe a certain percentage of all traffic within the Tor network, e.g., by adding fake nodes or by assuming control over existing nodes.

Definition 8 (Monetary Adversary). *A* monetary adversary *is a budget adversary $\mathcal{A}^{B}_{f_{\$}}$ that compromises Tor nodes with a monthly monetary maintenance and renting cost of at most B. For a set of providers $\mathcal{P}$ and a function* $\mathsf{price}\colon \mathcal{P} \times \mathbb{N} \to \mathbb{N}$ *that assigns a price for each provider and offered bandwidth, we have $f_{\$}(n) :=$* $\mathsf{price}(n.\texttt{provider}, n.\texttt{BW})$ *for $n \in \mathcal{N}$.*

Monetary adversaries reflect adversaries with a limited budget for the operational cost of running Tor nodes.

2.3 Anonymity Impact of a Budget Adversary

We now combine our formalization of observations and of their anonymity impact from Section 2.1 with our definition of a budget adversary. Thereby, we compute the anonymity impact of a budget adversary on Tor's path selection algorithm for each of the three considered anonymity notions.

The anonymity impact of all observations can be defined on a per-node basis if we slightly over-approximate the impact possible other adversarial nodes can have on the observations (or on the lack of observations). This so-called *indirect* impact captures the impact of a compromised node on the (lack of) observations made by other Tor entities, i.e., in addition to learning which compromised Tor nodes were used within a circuit, the adversary also learns which compromised Tor nodes were *not* used in the circuit, and can hence draw corresponding conclusions.

To define the indirect impact $\mathbf{Impact}^{\mathsf{ind}}_{X}$ for anonymity notion α_X, we use the following notation: (i) we write $N \overset{B,f}{\subseteq} \mathcal{N}$ instead of $N \subseteq \mathcal{N} s.t. \sum_{n' \in N} f(n') \leq B$; (ii) we write $\mathbf{CP}^{ab}_{n_g,n_m,n_x}$ as a shortcut for $\Pr\left[(\mathrm{S}_a, n_g, n_m, n_x, \mathrm{R}_b) \leftarrow \mathsf{ps}(\mathrm{S}_a, \mathrm{R}_b)\right]$; and (iii) we write $\mathbf{nCP}^{ab}_{n}$ instead of $\sum_{n',n'' \in \mathcal{N}} \mathbf{CP}^{ab}_{n,n',n''} + \mathbf{CP}^{ab}_{n',n,n''} + \mathbf{CP}^{ab}_{n',n'',n}$ to denote the probability that a node n is used within a circuit. Using these notions, we define six helper functions for computing $\mathbf{Impact}^{\mathsf{ind}}_{X}$ in Figure 3: $\mathbf{Impact}^{(ab,cd)}_{\mathsf{indirect}}$ defines the indirect impact that one compromised node has on the observations of another compromised node. $\mathbf{Impact}_{\mathsf{Rec1}}$ and $\mathbf{Impact}_{\mathsf{Rec2}}$ describe the impact of compromised nodes on observations made by a malicious recipient (used for sender anonymity) – the first notion considers the impact that any individual (compromised) node may have on the (lack of) observations of a malicious recipient, the second one bounds the maximal error in calculating the first one. $\mathbf{Impact}_{\mathsf{Sen1}}$ and $\mathbf{Impact}_{\mathsf{Sen2}}$ analogously describe the impact of compromised nodes on observations made by a malicious sender (used for recipient anonymity). Finally, $\mathbf{Impact}^{\mathsf{combined}}_{\mathrm{REL}}$ describes the slightly more complex direct impact of nodes on relationship anonymity; we in particular we need to consider the direct impact of two nodes per circuit for relationship anonymity. Based on these helper functions, the indirect impact $\mathbf{Impact}^{\mathsf{ind}}_{X}$ is defined in Figure 4.

We finally give the overall definition of *anonymity impact* and explain it in detail after the definition.

Definition 9 (Anonymity Impact)**.** *Let* $\mathrm{S}_0, \mathrm{S}_1$ *be two senders, let* $\mathrm{R}_0, \mathrm{R}_1$ *be two recipients, let* α_X *for* $X \in \{\mathrm{SA}, \mathrm{RA}, \mathrm{REL}\}$ *be an anonymity notion, and let* $\mathcal{A}_f^B$ *be a budget adversary. Then,* $\mathbf{Impact}_X(\mathcal{A}_f^B)$*, as defined in Figure 5, defines the* anonymity impact *of* $\mathcal{A}_f^B$ *for* α_X *and* $\mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1$.

The computation of the anonymity impact $\mathbf{Impact}_X(\mathcal{A}_f^B)$ depends on the considered anonymity notion α_X as follows.

Sender Anonymity. The impact of any budget adversary $\mathcal{A}_f^B$ on sender anonymity constitutes at most the aggregated observation impact of the optimal set of corrupted nodes together with the observation impact $\mathbf{Impact}_{\mathrm{SA}}^{\mathsf{obs}}(\mathrm{R}_0)$ of the malicious recipient R_0 (which is assumed for sender anonymity) and the indirect impact on sender anonymity, c.f., Equations (1) in Figures 4 and 5.

Recipient Anonymity. Analogously, the impact of any budget adversary $\mathcal{A}_f^B$ on recipient anonymity is, at most, the aggregated observation impact of the optimal set of corrupted nodes together with the observation impact $\mathbf{Impact}_{\mathrm{RA}}^{\mathsf{obs}}(\mathrm{S}_0)$ of the malicious sender S_0 (which is assumed for recipient anonymity) and the indirect impact on recipient anonymity, c.f., Equations (2) in Figures 4 and 5.

Relationship Anonymity. Analogously, the impact on any budget adversary $\mathcal{A}_f^B$ on relationship anonymity is, at most, the aggregated observation impact of the optimal set of corrupted nodes and the indirect impact on relationship anonymity, c.f., Equations (3) in Figures 4 and 5. In contrast to sender anonymity and recipient anonymity, the indirect anonymity impact is much more significant for relationship anonymity as, intuitively, both ends of the communication need to be deanonymized, c.f., $\mathsf{combined}_{\mathrm{REL}}$ in Figure 3.

Precision of our Calculation. For all circuit observations made by individual nodes, pairs of nodes, and sender or recipient, our calculation of $\mathbf{Impact}_{\mathrm{SA}}^{\mathsf{obs}}(N)$, $\mathbf{Impact}_{\mathrm{RA}}^{\mathsf{obs}}(N)$ and $\mathbf{Impact}_{\mathrm{REL}}^{\mathsf{obs}}(N)$ precisely captures the anonymity impact for the respective notion. However, when we aggregate the impact of individual nodes in $\mathbf{Impact}_X$ in order to derive our overall bounds for the anonymity impact of a budget adversary, we might count observations made for the same circuit more than once, therefore over-approximating the impact of the individual observations. Moreover, our bound on the (indirect) impact of nodes soundly overestimates the impact. We decided to accept this slight over-approximation for reasons of performance and scalability, as it allows us to compute bounds for budget adversaries based on each node individually; otherwise, we would have to combine all possible observations of all subsets of the set of nodes that fall within the budget. Furthermore, we implicitly assume that an adversary can mount traffic correlation attacks with perfect accuracy, i.e., whenever it observes traffic at two different points in the Tor network, we assume that the adversary can determine if this traffic belongs to the same Tor circuit. This assumption is motivated by the high accuracy achieved by recent work on traffic correlation attacks [20, 21, 24]; yet, it still constitutes an over-approximation.

3 Theoretical Underpinning

We now provide a rigorous semantics for the concepts that we informally used in the previous section, such as anonymity notions and the adversary's advantage. To this end, we cast all required formalizations in the AnoA framework [8], a framework for computing quantitative bounds for anonymous communication systems. By means of this embedding into AnoA, we show that $\mathbf{Impact}_X(\mathcal{A}_f^B)$, as defined in Definition 9, computes a bound for the notion of adversary's advantage in the AnoA framework for budget adversaries $\mathcal{A}_f^B$. In addition, we show the secure compositionality of budget adversaries in AnoA, which might be of independent interest.

3.1 Game-based Anonymity in AnoA

Anonymity Notions. The formalization of the three anonymity notions in AnoA closely follows the informal description that we gave in Section 2.1 as a challenge-response game, in which the adversary has to distinguish two scenarios. Formally, an *anonymity notion* is a function α that receives as inputs two senders: S_0 and S_1, two recipients: R_0 and R_1, and a so-called challenge bit b. It then selects one sender and one recipient, based on the challenge bit and the considered anonymity notion. For relationship anonymity, this selection is probabilistic.

Sender anonymity α_{SA}. The sender anonymity function α_{SA} selects the sender according to the challenge bit and

$$\begin{aligned}
\mathbf{Impact}^{(ab,cd)}_{\mathsf{indirect}}(n_g, n_x) &:= \sum_{n_m \in \mathcal{N}} \phi(\mathbf{CP}^{ab}_{n_g,n_m,n_x}, \mathbf{CP}^{cd}_{n_g,n_m,n_x}) \\
\mathbf{Impact}_{\mathsf{Rec1}}(n) &:= \sum_{n_x \in \mathcal{N}} \phi(\sum_{n' \in \mathcal{N}} (\mathbf{CP}^{10}_{n,n',n_x} + \mathbf{CP}^{10}_{n',n,n_x}), \sum_{n' \in \mathcal{N}} (\mathbf{CP}^{00}_{n,n',n_x} + \mathbf{CP}^{00}_{n',n,n_x})) \\
\mathbf{Impact}_{\mathsf{Rec2}}(n, n') &:= \sum_{n_x \in \mathcal{N}} \phi(\mathbf{CP}^{00}_{n,n',n_x} + \mathbf{CP}^{00}_{n',n,n_x}, \mathbf{CP}^{10}_{n,n',n_x} + \mathbf{CP}^{10}_{n',n,n_x}) \\
\mathbf{Impact}_{\mathsf{Sen1}}(n) &:= \sum_{n_g \in \mathcal{N}} \phi(\sum_{n' \in \mathcal{N}} (\mathbf{CP}^{01}_{n_g,n,n'} + \mathbf{CP}^{01}_{n_g,n',n}), \sum_{n' \in \mathcal{N}} (\mathbf{CP}^{00}_{n_g,n,n'} + \mathbf{CP}^{00}_{n_g,n',n})) \\
\mathbf{Impact}_{\mathsf{Sen2}}(n, n') &:= \sum_{n_g \in \mathcal{N}} \phi(\mathbf{CP}^{00}_{n_g,n,n'} + \mathbf{CP}^{00}_{n_g,n',n}, \mathbf{CP}^{01}_{n_g,n,n'} + \mathbf{CP}^{01}_{n_g,n',n}) \\
\mathbf{Impact}^{\mathsf{combined}}_{\mathrm{REL}}(n) &:= \mathbf{Impact}^{\mathsf{obs}}_{\mathrm{REL}}(n) + \max_{\substack{K \subseteq \mathcal{N} \setminus \{n\} \ s.t. \\ \sum_{m \in K} f(m) \leq B - f(n)}} \left(\sum_{m \in K} \mathbf{Impact}^{\mathsf{obs}}_{\mathrm{REL}}(\{n, m\}) \right).
\end{aligned}$$

Fig. 3. Notation for the indirect impact of nodes, as used in Figure 4

$$\begin{aligned}
&(1)\ \mathbf{Impact}^{\mathsf{ind}}_{\mathrm{SA}}(n, \mathcal{A}^B_f) := \mathbf{Impact}_{\mathsf{Rec1}}(n) + \max_{N \overset{B-f(n),f}{\subseteq} \mathcal{N}} \sum_{n' \in N} \left(\mathbf{Impact}^{(10,00)}_{\mathsf{indirect}} \left(n, n'\right) + \mathbf{Impact}_{\mathsf{Rec2}(n,n')} \right). \\
&(2)\ \mathbf{Impact}^{\mathsf{ind}}_{\mathrm{RA}}(n, \mathcal{A}^B_f) := \mathbf{Impact}_{\mathsf{Sen1}}(n) + \max_{N \overset{B-f(n),f}{\subseteq} \mathcal{N}} \sum_{n' \in N} \left(\mathbf{Impact}^{(01,00)}_{\mathsf{indirect}} \left(n', n\right) + \mathbf{Impact}_{\mathsf{Sen2}(n,n')} \right). \\
&(3)\ \mathbf{Impact}^{\mathsf{ind}}_{\mathrm{REL}}(n, \mathcal{A}^B_f) := \max_{K \overset{B-f(n),f}{\subseteq} \mathcal{N} \setminus \{n\}} \left(\frac{1}{2} \left(\mathbf{Impact}^{(01,00)}_{\mathsf{indirect}} \left(n, n'\right) + \mathbf{Impact}^{(10,11)}_{\mathsf{indirect}} \left(n, n'\right) \right.\right. \\
&\qquad\qquad \left.\left. + \mathbf{Impact}^{(01,11)}_{\mathsf{indirect}} \left(n', n\right) + \mathbf{Impact}^{(10,00)}_{\mathsf{indirect}} \left(n', n\right) \right)\right)
\end{aligned}$$

Fig. 4. Definition of indirect impact (for Definition 9)

always considers the same recipient R_0:

$$\alpha_{\mathrm{SA}}(\mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1, b) := (\mathrm{S}_b, \mathrm{R}_0).$$

Recipient anonymity α_{RA}. The recipient anonymity function α_{RA} selects the recipient according to the challenge bit and always considers the same sender S_0:

$$\alpha_{\mathrm{RA}}(\mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1, b) := (\mathrm{S}_0, \mathrm{R}_b).$$

Relationship anonymity α_{REL}. The relationship anonymity function α_{REL} selects one of the four possible sender-recipient combinations as follows: if $b = 0$, the function randomly selects one of the two pairs $(\mathrm{S}_0, \mathrm{R}_0)$ or $(\mathrm{S}_1, \mathrm{R}_1)$; if $b = 1$, it randomly selects between $(\mathrm{S}_0, \mathrm{R}_1)$ and $(\mathrm{S}_1, \mathrm{R}_0)$. In short, we obtain

$$\alpha_{\mathrm{RA}}(\mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1, b) := (\mathrm{S}_{b'}, \mathrm{R}_{b \oplus b'}); b' \leftarrow_R \{0, 1\}.$$

Game-based anonymity definition. The definition of the AnoA challenger is the final building block for the definition of the adversary's advantage in AnoA as a challenge-response game. In the AnoA framework, the challenger receives as input an anonymity notion α, a bound on the permitted challenge-messages (see-below), two senders, two recipients, and the challenge bit. It then simulates the Tor protocol for the sender-recipient scenario selected by α. The adversary interacts with the challenger in order to determine which scenario is being simulated. The adversary knows all inputs to the challenger up to an uncertainty of one bit (the challenge bit b). We now describe the challenger in detail.

The AnoA challenger. The challenger $\mathrm{C}\textsc{h}$ is defined in Figure 6. As described above, it expects as inputs the anonymity notion α, a bound γ on the permitted challenge-messages, two senders S_0, S_1, two recipients R_0, R_1, and the challenge bit b. The challenger initially waits for a set $N \subseteq \mathcal{N} \cup \{\mathrm{S}_0, \mathrm{R}_0\}$ of compromised Tor nodes (and for casting the different notions: respective sender S_0 and recipient R_0 of the challenge circuits, see below). The challenger first removes illegitimate corruption requests: S_0 is removed from N

$$(1)\ \mathbf{Impact}_{\mathrm{SA}}(\mathcal{A}_f^B) := \mathbf{Impact}^{\mathsf{obs}}_{\mathrm{SA}}(\mathrm{R}_0) + \max_{N \overset{B,f}{\subseteq}} \sum_{n \in N} \Big(\mathbf{Impact}^{\mathsf{obs}}_{\mathrm{SA}}(n) + \mathbf{Impact}^{\mathsf{ind}}_{\mathrm{SA}}(n, \mathcal{A}_f^B) \Big).$$

$$(2)\ \mathbf{Impact}_{\mathrm{RA}}(\mathcal{A}_f^B) := \mathbf{Impact}^{\mathsf{obs}}_{\mathrm{RA}}(\mathrm{S}_0) + \max_{N \overset{B,f}{\subseteq}} \sum_{n \in N} \Big(\mathbf{Impact}^{\mathsf{obs}}_{\mathrm{RA}}(n) + \mathbf{Impact}^{\mathsf{ind}}_{\mathrm{RA}}(n, \mathcal{A}_f^B) \Big).$$

$$(3)\ \mathbf{Impact}_{\mathrm{REL}}(\mathcal{A}_f^B) := \max_{N \overset{B,f}{\subseteq} \mathcal{N}} \sum_{n \in N} \Big(\mathbf{Impact}^{\mathsf{combined}}_{\mathrm{REL}}(n) + \mathbf{Impact}^{\mathsf{ind}}_{\mathrm{REL}}(n, \mathcal{A}_f^B) \Big)$$

Fig. 5. Definition of $\mathbf{Impact}_X(\mathcal{A}_f^B)$, in order to define anonymity impact (Definition 9)

for sender anonymity, R_0 is removed from N for recipient anonymity, and both S_0 and R_0 are removed from N for relationship anonymity, which reflects the respective scenarios. Then, it accepts two types of messages from the adversary: *challenge-messages*, denoted as $(\mathsf{challenge}, m)$ in Figure 6, that trigger that a challenge message is sent, and *input-messages*, denoted as $(\mathsf{input}, \mathrm{S}, m, \mathrm{R})$ in Figure 6, that send additional messages m between senders S and recipients R:

- *Challenge-messages:* Upon receiving a message $(\mathsf{challenge}, m)$, the challenger increases Ψ and only proceeds if Ψ is still less than or equal to γ. It then computes the anonymity notion α on $(\mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1)$ and the challenge bit b and obtains a sender-recipient pair $(\mathrm{S}^*, \mathrm{R}^*) \in \{\mathrm{S}_0, \mathrm{S}_1\} \times \{\mathrm{R}_0, \mathrm{R}_1\}$. The challenger then simulates the Tor protocol by creating a new Tor circuit (n_g, n_m, n_x) from sender S^* to recipient R^*, and then sends the message m from S^* to R^* using Tor. We abbreviate this using the subroutine $\mathbf{SimulateTor}(\mathrm{S}^*, m, \mathrm{R}^*)$ in Figure 6. Whenever a node n involved in the constructed circuit is considered corrupted, i.e., $n \in N$, or if $\mathrm{S}^* = \mathrm{S}_0 \in N$ or $\mathrm{R}^* = \mathrm{R}_0 \in N$, then the adversary is given the transcript of this communication, i.e., the messages n sent and received in this circuit.
- *Input-messages:* Upon receiving a message $(\mathsf{input}, \mathrm{S}, m, \mathrm{R})$, the challenger calls the subroutine $\mathbf{SimulateTor}(\mathrm{S}, m, \mathrm{R})$, as described above. Input-messages, hence, capture additional information the adversary may have about the communication contents in the Tor network.

We now define the *reduction of anonymity* for α as the adversary's advantage in this game.

Definition 10 (Reduction of anonymity; advantage). *Let α be an anonymity notion, $\gamma \in \mathbb{N}$, $\mathrm{S}_0, \mathrm{S}_1$ two senders, and $\mathrm{R}_0, \mathrm{R}_1$ two recipients. Then, the* adversary's advantage *of an adversary $\mathcal{A}$ for these parameters is at most δ, with $0 \le \delta \le 1$, if for all sufficiently large $\eta \in \mathbb{N}$, we have*

$$\begin{aligned} &\Pr\left[0 = \langle \mathcal{A}(1^\eta) || \textsc{Ch}(\alpha, \gamma, \mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1, 0) \rangle\right] \\ \le\ &\Pr\left[0 = \langle \mathcal{A}(1^\eta) || \textsc{Ch}(\alpha, \gamma, \mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1, 1) \rangle\right] + \delta. \end{aligned}$$

We say that Tor exhibits a reduction of anonymity *of at most δ under γ challenges (formally: Tor is (δ, γ)-$\mathsf{IND\text{-}ANO}$) for these parameters $\alpha, \mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1$ and a class A of adversaries if the adversary's advantage of all probabilistic polynomial-time adversaries $\mathcal{A} \in A$ is at most δ.*

This definition captures an eavesdropping adversary that corrupts a fixed set of nodes before it starts observing the network. In particular, the adversary cannot adaptively decide which nodes to compromise.[2]

Relation to Entropy-based Anonymity Notions. Our indistinguishability-based notion reasons about the cryptographic implementation of Tor. For such cryptographic systems with their computational security guarantees, entropy-based notions, including notions that define the *effective size* of an anonymity set [13, 23], are not directly applicable. A relation between entropy-based notions and cryptographic notions might be possible along the lines of [11] that establishes a tight correspondence between the information-theoretic capacity of channels, their abstract description and finally their cryptographic instantiations. We plan to investigate this approach in the context of more comprehensive systems such as Tor in future work.

[2] The corresponding definition in the AnoA paper [8] additionally considers a multiplicative advantage e^{ε}. We have set this to 1 in this paper, such that δ directly corresponds to the reduction of anonymity. Moreover, AnoA considers arbitrary probabilistic, polynomial-time Turing machines and, for technical reasons, subsequently restricts them with wrapper machines (so-called adversary classes). For the sake of presentation in our specific setting, we did not introduce the lengthy description of adversary classes, but instead restricted the adversary in the core definition and adjusted the challenger accordingly.

```
ANOA Challenger $\textsc{Ch}(\alpha, \gamma, \mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1, b)$
Initial message (corruption setting)
  Receive $N \subseteq \mathcal{N} \cup \{\mathrm{S}_0, \mathrm{R}_0\}$ as input.
  if $\alpha = \alpha_{\mathrm{SA}}$, let $N := N \setminus \{\mathrm{S}_0\}$.
  if $\alpha = \alpha_{\mathrm{RA}}$, let $N := N \setminus \{\mathrm{R}_0\}$.
  if $\alpha = \alpha_{\mathrm{REL}}$, let $N := N \setminus \{\mathrm{S}_0, \mathrm{R}_0\}$.
  $\Psi := 0$.
Upon message $(\mathsf{input}, \mathrm{S}, m, \mathrm{R})$
  Run SimulateTor$(\mathrm{S}, m, \mathrm{R})$.
Upon message $(\mathsf{challenge}, m)$
  $\Psi := \Psi + 1$.
  if $\Psi \leq \gamma$ then
    Compute $(\mathrm{S}^*, \mathrm{R}^*) \leftarrow \alpha(\mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1, b)$
    Run $\mathsf{SimulateTor}(\mathrm{S}^*, m, \mathrm{R}^*)$
  else abort the game.
Subroutine SimulateTor$(\mathrm{S}, m, \mathrm{R})$
  Simulate the Tor protocol:
    S builds a fresh Tor circuit $C$, yielding $(n_g, n_m, n_x)$.
    S sends $m$ to R via the circuit $C$.
    for each $n \in \{\mathrm{S}, n_g, n_m, n_x, \mathrm{R}\} \cap N$ do
      Output the transcript of $n$ in $C$.
```

Fig. 6. Definition of the AnoA Challenger

3.2 Budget Adversaries in AnoA

We now cast the notion of a budget adversary in AnoA. Intuitively, an adversary is a budget adversary if the set of corrupted nodes that it sends to the challenger conforms to its budget restrictions.[3]

Definition 11 (AnoA budget adversary). *Consider a function $f \colon \mathcal{N} \to \mathbb{N}$ and a budget $B \in \mathbb{N}$. Then a probabilistic polynomial-time adversary $\mathcal{A}_f^B$ is an* AnoA budget adversary *for f and B, if $\mathcal{A}_f^B$ if for all possible outputs N to the AnoA challenger in its first message (denoting the corrupted nodes), we have that $\sum_{n \in M} f(n) \leq B$ for $M := N \setminus \{\mathrm{S}_0, \mathrm{R}_0\}$. We let A_f^B denote the class of all budget adversaries for f and B.*

All instantiations of budget adversaries defined in Section 2.2 can be cast in AnoA in an analogous manner.

As a result that we consider to be of independent interest, we show that anonymity guarantees for *individual* challenges against an AnoA budget adversary entail anonymity guarantees for an *arbitrary* (fixed) amount of challenges (for different, but related parameters). Formally, AnoA budget adversaries A_f^B are *composable* for every budget B and every cost function f.

Theorem 1 (Composition). *If Tor is $(\delta, 1)$-IND-ANO for an anonymity notion α, two senders $\mathrm{S}_0, \mathrm{S}_1$, two recipients $\mathrm{R}_0, \mathrm{R}_1$, and the class of budget adversaries A_f^B, then, Tor is also $(\gamma \cdot \delta, \gamma)$-IND-ANO for α, $\mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1$ and A_f^B, for every $\gamma \in \mathbb{N}$.*

For space reasons, we postpone this and upcoming proofs to Appendix A.

3.3 Correctness of Impact_X bounds

We now show that $\mathbf{Impact}_X(\mathcal{A}_f^B)$, as defined in Definition 9, closely corresponds to the notion of adversary's advantage in the AnoA framework for budget adversaries $\mathcal{A}_f^B$, thereby establishing the output of $\mathbf{Impact}_X$ as accurate bounds for Tor against such adversaries.

We first show that our calculation of observation impact exactly corresponds to the *optimal advantage* of any adversary that makes those observations, provided that the adversary only corrupts one node, that it only sends a single challenge-message, and that we consider an idealization of cryptography. We call an advantage of δ *optimal* for a class of adversaries A if the adversary's advantage for sufficiently large η is at most δ for all probabilistic polynomial-time adversaries $\mathcal{A} \in A$ and if there exists an adversary $\mathcal{A} \in A$ that achieves this advantage, i.e., the less-or-equal in Definition 10 is replaced by equality for $\mathcal{A}$ for sufficiently large η.

For a sender S_0, a recipient R_0, and $X \in \{SA, RA, REL\}$, let $A_{\mathrm{S}_0,\mathrm{R}_0,X}$ be the class of probabilistic, polynomial-time adversaries that corrupt precisely one node and that only send one challenge-message to $\textsc{Ch}$. More precisely, $A_{\mathrm{S}_0,\mathrm{R}_0,X}$ sends a singleton $\{n\} \subseteq \mathcal{N} \cup \{\mathrm{S}_0, \mathrm{R}_0\}$ to $\textsc{Ch}$ as the corrupted node n, and we have that $n \neq \mathrm{S}_0$ if $X \in \{\mathrm{SA}, \mathrm{REL}\}$ and that $n \neq \mathrm{R}_0$ if $X \in \{\mathrm{RA}, \mathrm{REL}\}$.

To define the idealization of cryptography, we define an idealized AnoA challenger $\textsc{Ch}^*$. $\textsc{Ch}^*$ is defined exactly as $\textsc{Ch}$, with the only difference that in the subroutine **SimulateTor**, where $\textsc{Ch}$ sends the transcript of

[3] An alternative, yet slightly more technical way of defining budget adversaries in AnoA is by using the concept of adversary classes provided by AnoA: in this case, a budget adversary would be a wrapper machine that internally runs an arbitrary adversary as a black box, but ensures that only corruption requests N are being forwarded to the AnoA challenger that satisfy $\sum_{n \in M} f(n) \leq B$ for $M := N \setminus \{\mathrm{S}_0, \mathrm{R}_0\}$. We opted for the technically simpler definition in this paper.

messages sent and received to $\mathcal{A}$, $\textsc{Ch}^*$ only sends n and its predecessor and successor to $\mathcal{A}$. This models that the adversary cannot gain information about the content of encrypted messages, but that it can still determine at which point in the challenge circuit it makes an observation, and then derive predecessor and successor. The only exception is that if the observation is made at the exit node or at the recipient, then the adversary would be able to see the message. However, since the adversary is allowed to choose the message in the interaction with the challenger anyway, observing it does not reveal additional information.

Lemma 1. *For every anonymity notion α_X with $X \in \{SA, RA, REL\}$, all senders $\mathrm{S}_0, \mathrm{S}_1$ and recipients $\mathrm{R}_0, \mathrm{R}_1$, the optimal advantage for $A_{\mathrm{S}_0,\mathrm{R}_0,X}$ is equal to $\mathbf{Impact}_X^{\mathsf{obs}}(n)$.*

Using Lemma 1 we can now show our main theorem.

Theorem 2 (Soundness)**.** *For every anonymity notion α_X with $X \in \{\mathrm{SA}, \mathrm{RA}, \mathrm{REL}\}$, all senders $\mathrm{S}_0, \mathrm{S}_1$ and recipients $\mathrm{R}_0, \mathrm{R}_1$, for every budget B and every cost function f, Tor is $(\delta, 1)$-IND-ANO for the class of budget adversaries A_f^B, where $\delta = \mathbf{Impact}_X(A_f^B)$, as calculated in Section 2.2, up to a negligible additive factor.*

4 Evaluation

In this section, we apply the computation proposed in Section 2 to recent Tor Metrics data [5] to quantify the anonymity impact of various budget adversaries. Each of these adversaries is evaluated for Tor's standard path selection algorithm (short: TorPS) and for several commonly considered variants including SelekTOR [22], DistribuTor [10], and LASTor [7]. Furthermore, we compute the anonymity impact of a path selection transition phase, i.e., a small number of pioneering users decide to use an alternative path selection algorithm, while the remaining users still run the original algorithm. We stress again that our evaluation assesses the anonymity impact of structural attacks only, without taking any potential countermeasures by the users or the Tor developers into account. Moreover, our evaluation only addresses the anonymity impact, and disregards a potentially detrimental performance impact of the respective path selection algorithm.

We structure the section as follows. We first briefly review the evaluated path selection algorithms. We then describe how we implemented the computation of $\mathbf{Impact}_X(A_f^B)$, how we selected senders and recipients for our analyses and which adversaries we considered in our evaluation. Finally, we present and discuss the corresponding results for the anonymity impact.

4.1 Evaluated Path Selection Algorithms

The computation of $\mathbf{Impact}_X$ in Section 2 relies on the probability distribution $\mathsf{ps}(\mathrm{S}_i, \mathrm{R}_j)$ over Tor circuits that is induced by the considered path selection algorithm for sender S_i and recipient R_j. For our evaluation, we concretely instantiate this distribution using the Tor network consensus data. The dependence of Tor's path selection algorithm on a multitude of parameters (e.g., individual flags and weights of Tor nodes, family relations, TCP ports required for a connection, parameters selected by senders, etc.) makes this a non-trivial task. In addition to Tor's default path selection algorithm, several variants have been proposed that strive to improve performance or anonymity under specific assumptions. All of these variants are characterized by the different probability distributions over Tor circuits they induce, and they can, hence, be evaluated by means of $\mathbf{Impact}_X$ as well.

TorPS – Tor's Standard Path Selection. We utilize MATor [10] for computing the distribution of TorPS. TorPS randomly selects nodes based on their flags in the Tor consensus (e.g., only nodes with the *guard* flag can become guard nodes, nodes with the flag *bad-exit* cannot be used as exit nodes, etc.) and, in addition, for the exit node based on whether the ports required by the user are offered by the Tor node. The path selection weights this random choice with the weight in the Tor consensus. If the sender has created at least one circuit, the guard node selected in that circuit is used as the guard node in all subsequent circuits as well. We refer to Tor's specification [6] and to MATor [10] for a more detailed description.

UniformTor. Many existing works abstract Tor's actual path selection as a uniform path selection algorithm. In this variant of Tor, all (eligible) entry, middle, and exit nodes are chosen with the same weight.

SelekTOR. SelekTOR [22] restricts the Tor client to always select an exit node from a specific country, e.g., in order to bypass geo-restrictions of websites and services. SelekTOR only differs from TorPS in that the weights of all Tor exit nodes outside the considered country are set to zero. In our evaluation, we consider a SelekTOR configuration with exit nodes in the US.

DistribuTor. DistribuTor [10] aims to mitigate the anonymity impact of Tor nodes with very large bandwidths by distributing the usage of guard nodes and exit nodes to the greatest possible extent. To this end, DistribuTor modifies the node weights so that nodes with a very large bandwidth are mostly used as middle nodes. Never used as middle nodes are those nodes with the guard or exit flag that have a low bandwidth.

LASTor. LASTor [7] groups Tor nodes together into so-called clusters based on their physical location (latitude and longitude), which it infers by their IP address via GeoIP. LASTor first selects a guard cluster, a middle cluster, and an exit cluster. In this weighted random selection, LASTor weights the clusters inverse to the distance of the path over them, where this path starts with the sender and ends with the recipient, thereby reducing the expected physical distance. After selecting clusters, LASTor selects a node from each cluster uniformly at random.

4.2 Implementation

We have implemented the computation of $\mathbf{Impact}_X$ as an extension of the MATor [10] tool. MATor already takes care of computing the probability distribution over Tor circuits for a given Tor network consensus and the respective server descriptors. Hence, we added the calculations from Section 2 for any given budget adversary $A_f^B(\cdot)$, the desired anonymity notion α_X, and concrete senders $\mathrm{S}_0, \mathrm{S}_1$ and recipients $\mathrm{R}_0, \mathrm{R}_1$. To this end, we first compute the individual impacts $\mathbf{Impact}^{\mathsf{obs}}$ for all observations of individual nodes, pairs of nodes and – depending on the anonymity notion – relevant end points. Leveraging the computation from $\mathbf{Impact}_X^{\mathsf{obs}}$ to $\mathbf{Impact}_X$ requires us to solve the underlying integer maximization problems, e.g., to determine $N \subseteq \mathcal{N}$ such that $\sum_{n \in N} f(n) \leq B$ becomes maximal. While this problem is known to be NP-hard, we can solve it using a simple dynamic programming algorithm since the number of Tor nodes, and hence the size of the considered instances, is sufficiently small. The source code of our implementation is available [2].

4.3 Senders and Recipients

Recall that our computations are with respect to specific senders $\mathrm{S}_0, \mathrm{S}_1$ and recipients $\mathrm{R}_0, \mathrm{R}_1$. For the sake of evaluation, we hence consider concrete users in the following: the IP addresses from the affiliations of the PC chairs of PETS2015 and PETS2016. The first user, S_0, establishes a Tor circuit from Drexel University in Philadelphia; the second user, S_1, connects from Indiana University in Bloomington. As possible destinations, we have selected TU Darmstadt as R_0 and KU Leuven as R_1. For both destinations, we only required the HTTPS port 443 as the by far most widely used port for Tor connections (Tor is mainly used via the Tor-Browser bundle, which includes HTTPS-Everywhere).

4.4 Evaluated Adversary Classes

We consider the following six instances of budget adversaries in our analysis. We evaluate the first four instances for all considered path selection algorithms, whereas the last two instances are specific for TorPS.

***k*-collusion adversary.** We evaluate the k-collusion adversary for up to 25 compromised nodes, i.e., for a budget B ranging from 0 to 25.

Bandwidth adversary. We evaluate the bandwidth adversary for a budget B ranging from 1 MB/s to 10 GB/s.

Geographic adversary. We evaluate several adversaries that compromise all nodes in a given country or set of countries. We consider the four top countries according to offered Tor bandwidth: Germany, France, the Netherlands and the US. Moreover, we consider a collaboration of all countries of the European Union (abbreviated EU) and a collaboration of the US, New Zealand, Canada, the United Kingdom, and Australia (the so-called Five Eyes, abbreviated FVEY).

Monetary adversary. We evaluate a monetary adversary with a monthly budget B in US dollars, ranging from 10^3 to 10^8 US dollars. Recall that the cost function $f_\$$ assigns each node its monthly cost, depending on a price function $\mathsf{price}(n.\mathsf{provider}, n.\mathtt{BW})$. We instantiate price for the 8 largest providers hosting Tor nodes (Amazon, DigitalOcean, Hetzner, LeaseWeb, myLoc, Online, OVH, and STRATO), accounting for approximately $\frac{1}{3}$ of Tor bandwidth as follows. For each provider P in this list we set $\mathsf{price}(P, \mathtt{BW})$ to the cost of the cheapest server offered by this provider that has at least a bandwidth of $\mathtt{BW}$. For all remaining nodes (that are not hosted by these providers), $\mathsf{price}(\cdot)$ assigns the average consumer price per bandwidth, depending on the node's country, taken from Ookla's NetIndex [1] per country.

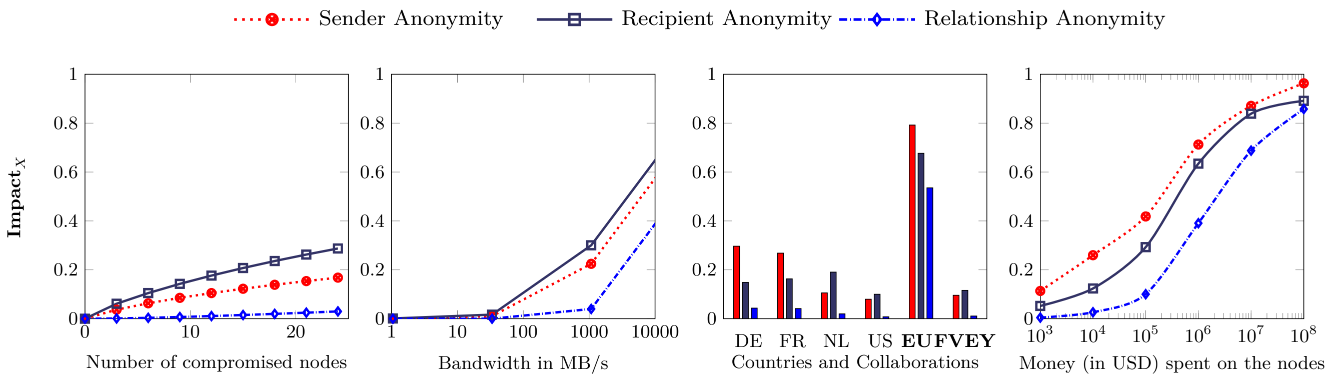


Fig. 7. **Impact$_X$** for different adversaries classes (from left to right): k-collusion, bandwidth, geographic, and monetary.

Vulnerable Tor Versions (TorPS only): We evaluate a predicate adversary for a critical Tor software update. The recently released update 0.2.6.10 solves many stability issues. The Tor blog recommends that every Tor node running an older version, especially an older version of 0.2.6 should update [4]. For presenting an example analysis of a Tor version predicate, we assume that there was a vulnerability in all Tor versions prior to the 0.2.6 branch and that the Vulnerable Tor adversary compromises all Tor nodes that run a Tor version between 0.2.6 and 0.2.6.10.

Guard Discovery Adversary (TorPS only): Recently, Tor has implemented a new strategy for selecting guard nodes called "One Fast Guard for Life" ([14]) that aims at improving longtime sender anonymity (and relationship anonymity to some extent). A sender selects a guard node once and uses it continuously over a period of 9 months to mitigate the danger imposed by frequently selecting fresh guard nodes and, moreover, to mitigate the danger imposed by selecting a recognizable set of guard nodes. We evaluate the anonymity impact of the first four aforementioned adversaries on the anonymity of guard nodes, effectively measuring their success to perform guard discovery attacks. Strictly speaking, this adversary does not constitute a single budget adversary, but a specific setting that is parametric in a considered budget adversary.

4.5 Results

Unless stated otherwise in a specific experiment, we used the following data in our evaluation: Our evaluation was conducted on Tor network consensus data over the course of one year (August 2014-July 2015), where we calculated the anonymity impact of the considered adversaries on four consensus data per day (at midnight, 6 a.m., noon and 6 p.m.). Additionally, we conducted extensive evaluations over the course of one year for the first four adversaries considered in Section 4.4, with the following budget choices: 10 compromised nodes for k-collusion adversary, 1 GB/s for the bandwidth adversary, Five Eyes for the geographic adversary, and 100,000 USD/mo monthly budget for the monetary adversary. The results – for sender, recipient and relationship anonymity – are depicted in Figure 12 in Appendix A, where we averaged the results per day to minimize day-time dependencies. For all other graphs we averaged all results to minimize any short-time impacts.

Remark: Please note that all of our evaluations consider the *worst-case* adversary for the respective class, i.e., we calculate and plot the maximal adversarial impact within this class to give an anonymity guarantee against this type of adversary. For example, the k-collusion adversary will compromise the k nodes with the highest impact for the considered anonymity notion and not just some subset of k nodes. Likewise, we consider the worst-case adversary for each notion separately, i.e., the adversary may compromise different nodes for each of the anonymity notion.

4.5.1 Evaluating Tor's Path Selection Algorithm

The results of our evaluation of TorPS are depicted in Figure 7.

Results – k-collusion adversary (left, Fig. 7). Our results confirm the (known) strong anonymity impact of a small number of high-bandwidth Tor nodes – a collusion of 10 Tor nodes results in a reduction of sender anonymity of 9.2%, a reduction of recipient anonymity of 15.4% and a reduction of relationship anonymity of

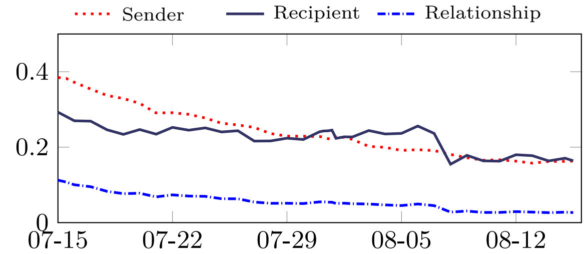


Fig. 8. Adversarial **Impact** for sender, recipient and relationship anonymity against the *Vulnerable Tor Version* adversary.

0.8%. The reduction of anonymity grows sub-linearly in k for sender anonymity and recipient anonymity, and more than linearly in k for relationship anonymity. A reduction of anonymity of $> 95\%$ amounts to 915 compromised nodes for sender anonymity, 330 compromised nodes for recipient anonymity, and 1230 compromised nodes for relationship anonymity.

Results – bandwidth adversary (2nd left, Fig. 7). The plot shows the reduction of anonymity on a logarithmic scale axis. An adversary compromising nodes with at most 1 GB/s of average bandwidth achieves a reduction of sender anonymity of 22.4%, a reduction of recipient anonymity of 30.0%, and a reduction of relationship anonymity of 3.8%. A reduction of anonymity of $> 95\%$ amounts to compromised bandwidth of 250 GB/s (for sender anonymity), 65 GB/s (for recipient anonymity), and 310 GB/s (for relationship anonymity).

Results – geographic adversary (2nd right, Fig. 7). Our results in particular show that no country on its own can successfully break relationship anonymity with a significant probability. However, their collaboration, in the case of the European Union, would be capable of deanonymize a significant amount of Tor traffic (anonymity impact of 53% for relationship anonymity), which significantly surpasses the advantage of the Five Eyes adversary (1.1%). We stress again that these results are specific to structural attacks against nodes, and do not take adversary-controlled network structure into account (such as monitoring traffic of domestic ISPs).

Results – monetary adversary (right, Fig. 7). An adversary running Tor nodes with a monthly cost of 100,000 USD at most reduces sender anonymity by 41.8%, recipient anonymity by 29.3% and relationship anonymity by 10.0%. The smaller reduction of recipient anonymity compared to sender anonymity stems from the fact that the prices of hosting guard nodes, on average, are significantly lower than the prices of hosting exit nodes. A reduction of anonymity of $> 95\%$ amounts to monthly costs of 8.75 Mio. USD (for sender anonymity), 27.5 Mio. USD (for recipient anonymity), and 40 Mio. USD (for relationship anonymity).

Vulnerable Tor Versions (Fig. 8). Our evaluation shows that on August 15th, i.e., one month after the release of the fix, the Tor version adversary still achieves a reduction of anonymity of 16.33% for sender anonymity, of 16.4% for recipient anonymity and of 2.67% for relationship anonymity. We refer to Figure 8 for a graph showing the reduction of anonymity of this adversary over the course of one month after the release of Tor version 0.2.6.10.

Anonymity Against Guard Discovery Attacks (Fig. 11). We selected a set of Tor consensus data (28 consensus data from July 23rd to July 29th 2015, taken each 6 hours) and from each of these consensus data we selected the top 25 guard nodes that share their /16 subnet with at least one exit node (which effectively affects around 45% of Tor guard nodes). Subsequently, we compared the anonymity impact of a budget adversary in distinguishing these guards, i.e., for all pairs of selected guard nodes (n_{g0}, n_{g1}) with $n_{g0} \neq n_{g1}$, we proceeded as follows: First, S_0 selects n_{g0} as guard node and S_1 selects n_{g1} as guard node. Second, we set the costs for compromising n_{g0} or n_{g1} to ∞ to disallow the adversary to compromise the respective guard nodes. Finally, we compute the sender anonymity impact.

Even without compromised nodes, a compromised recipient reduces sender anonymity by 2.12% on average. We furthermore evaluated our four aforementioned adversaries against the selected pairs of guard nodes. Our results, depicted in Figure 11 in Appendix A, show that the relationship between the adversarial advantage in the normal case and our guard discovery strategy differs strongly for the considered adversaries. For the k-collusion adversary for instance, with more than 8 compromised nodes, the adversary obtains a smaller anonymity impact in the guard-discovery scenario; in contrast to that, the difference in the anonymity impact of a monetary adversary becomes larger the more money it spends.

4.5.2 Alternative Path Selection Algorithms

The results of our evaluation of alternative path selection algorithms are depicted in Figure 9 – from top to bottom: sender anonymity, recipient anonymity, and relationship anonymity, showing the differences between the anonymity impacts of the alternative path selection algorithm and TorPS.

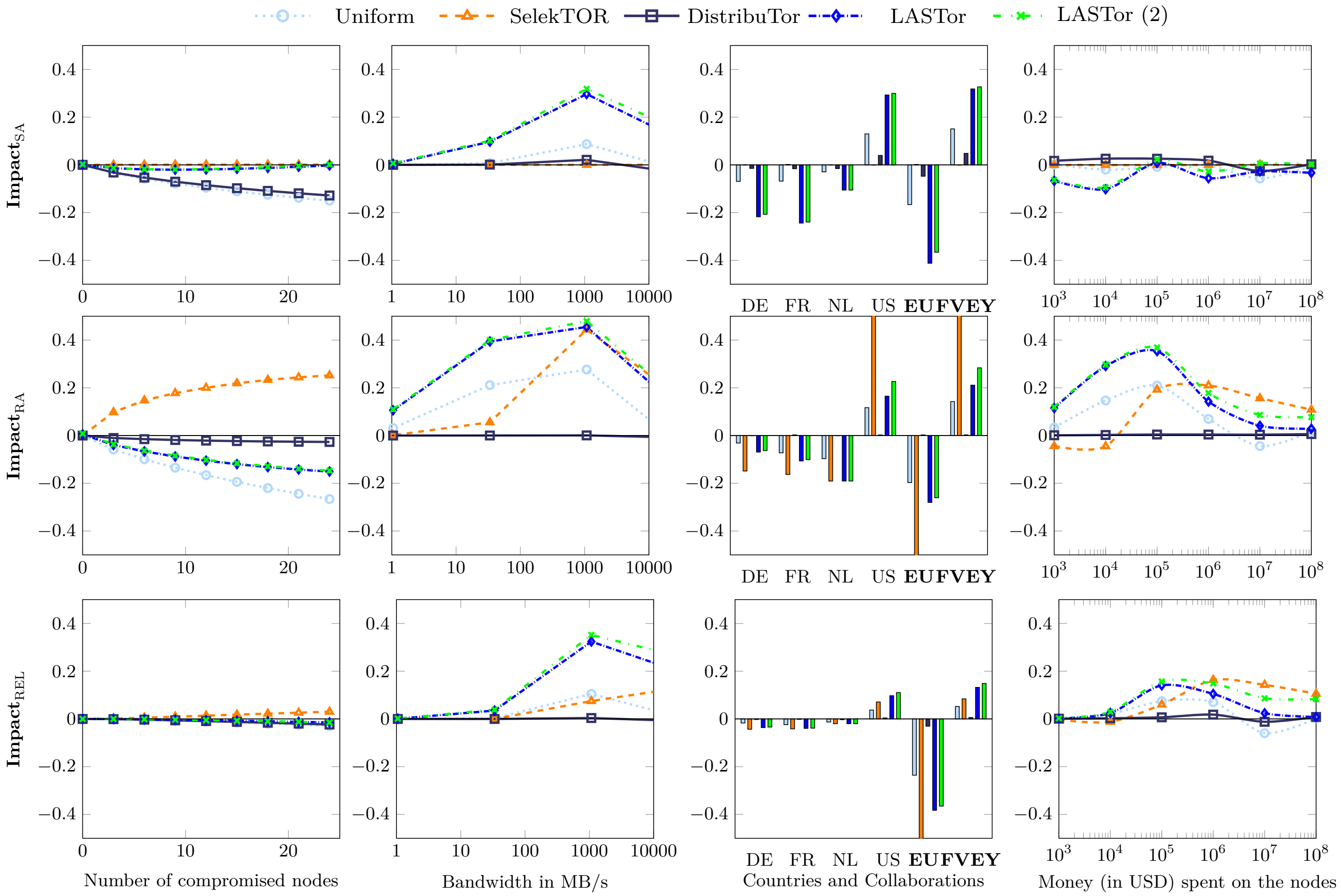


Fig. 9. The difference between $\textbf{Impact}_X$ for sender (top), recipient (middle) and relationship (bottom) anonymity of alternative path selection algorithms and TorPS. Cionsidered adversaries classes (from left to right): k-collusion, bandwidth, geographic, and monetary.

UniformTor. It is commonly believed that the uniform distribution over all nodes offers the highest degree of anonymity. Against k-collusion adversaries, this is certainly true. However, an adversary that corrupts a certain amount of bandwidth can corrupt a large number of low-bandwidth nodes even with a small budget. In case of a monetary adversary, 10 Mio. USD constitutes the break-even point at which the uniform path selection actually exhibits better anonymity guarantees (but at this point, anonymity has degenerated to a large extent anyway). The anomaly is caused by expensive nodes contributing a very small advantage in case of uniform path selection.

SelekTOR. As SelekTOR only restricts the choice for an exit node, sender anonymity is comparable to Tor, while relationship anonymity and in particular recipient anonymity suffers significantly, for essentially all considered adversaries. As expected, the geographic adversary that compromises all nodes within the US can break recipient anonymity with 100% probability (it always controls the exit node and can perform a traffic correlation).

DistribuTor. DistribuTor in particular modifies the selection of entry nodes by capping the possible weights of nodes at a certain point. Consequently, it achieves better sender anonymity and relationship anonymity guarantees against k-collusion adversaries (against which it has been designed), but does not exhibit a clear advantage against bandwidth adversaries. Since it uses modified weights (especially for entry nodes), its sender anonymity is, in comparison to Tor's sender anonymity, slightly less prone against European country adversaries, but more vulnerable against the US country adversary (since there are more smaller entry nodes in the US).

LASTor. Since our structural adversaries do not perform network-based attacks, an evaluation of LASTor, which is designed to counter network-based attacks, is slightly unfair. Still we gained interesting insights by our analysis. The uniform distribution of the node weights within LASTor's "geo-location buckets", leads to significantly better results against a k-collusion adversary. However, even a small amount of compromised bandwidth suffices for completely breaking anonymity, as even a small, compromised middle node can gain information about the location of both sender and recipient of a communication: entry and exit nodes have significantly different weights for different locations. Note that LASTor (as presented in [7]) additionally restricts circuits depending on whether traffic is expected to be routed through the same autonomous system twice. This additional restriction, however, only increases the advantage of a structural adversary.

To further evaluate LASTor, we also ran analyses swapping one recipient with one sender (TU Darmstadt with Indiana University). This is displayed as *LASTor (2)* in Figure 9. As expected, this more diverse selection of sender and recipient had a negative impact on the anonymity guarantees.

4.5.3 Transition Phase

In existing evaluations of alternative path selection algorithms, the major impeding anonymity factor that is typically omitted is the so-called transition phase, i.e., a small number of users is already using an alternative of Tor's path selection algorithm, whereas the vast majority still uses the standard one. Intuitively, there are not yet enough users with the alternative path selection algorithm to provide a sufficienly large anonymity set. Figure 10 depicts the adversary's advantage in such scenarios (for sender and relationship anonymity only, since these are the only two anonymity notions affected by this algorithmic transition).

Our analyses show that even adversaries that do not compromise any single Tor node have a tremendous advantage in distinguishing a user that relies on an alternative path selection algorithm from a regular Tor user: for SelekTOR, the adversary has an advantage of 92.89%, for LASTor an advantage of 85.87% and for the uniform path selection, the adversary has an advantage of 68.14%. In case of SelekTOR, this advantage arises mostly from the fact, that a normal user would choose a US exit node with only $\approx 7\%$ probability, whereas a SelekTOR user always uses such an exit node. For LASTor and the uniform path selection algorithm, circuits containing small nodes are chosen with a much higher probability in comparison to TorPS. The effect of the transition phase on DistribuTor is less drastic, but still noticeable. We attribute this result to the close similarity of DistribuTor and TorPS. However, the fact that the weights of entry nodes are heavily modified grants compromised middle nodes an advantage in distinguishing between a TorPS user and a DistribuTor user.

Mitigating the Risk of the Transition Phase. The high vulnerability of users that use a non-standard path selection algorithm indicates that a slow and voluntary transition from one algorithm to another might alienate the (few) users that migrate first and thus significantly weaken their anonymity. We think that as soon as a transition is necessary, the novel algorithm should be rolled out to all users at once, in order to shorten the transition phase. With this strategy, all users would intuitively remain in the same anonymity set.

5 Conclusion

In this paper, we have presented a rigorous methodology for quantifying the anonymity provided by Tor against a variety of structural attacks, i.e., adversaries that corrupt Tor nodes and thereby perform eavesdropping attacks to deanonymize Tor users. We have made the following two tangible contributions. First, we have provided the first algorithmic approach for soundly computing the anonymity impact of such structural attacks against Tor. We have devised formalizations of various instantiations of structural attacks against Tor, and we have subsequently proven that the computed anonymity impact for each of these adversaries constitutes a worst-case anonymity bound for the cryptographic realization of Tor, up to a negligible additive factor. We have furthermore shown that our approach is sequentially composable, which we consider to be of independent interest. Second, we have demonstrated the applicability of our approach to large-scale analyses by performing the to date largest rigorous anonymity evaluation of Tor's anonymity. Concretely, we have established worst-case anonymity bounds against various structural attacks for Tor and for alternative path selection algorithms such as DistribuTor, SelekTOR, and LASTor, yielding the first rigorous, quantitative anonymity comparison between these algorithms. We have moreover quantified the anonymity impact of a path selection transition phase, showing that a small number of pioneering users who decide to run a (potentially improved) alternative path selection algorithm first while the majority of users still runs the original algorithm, face severe risks of being deanonymized.

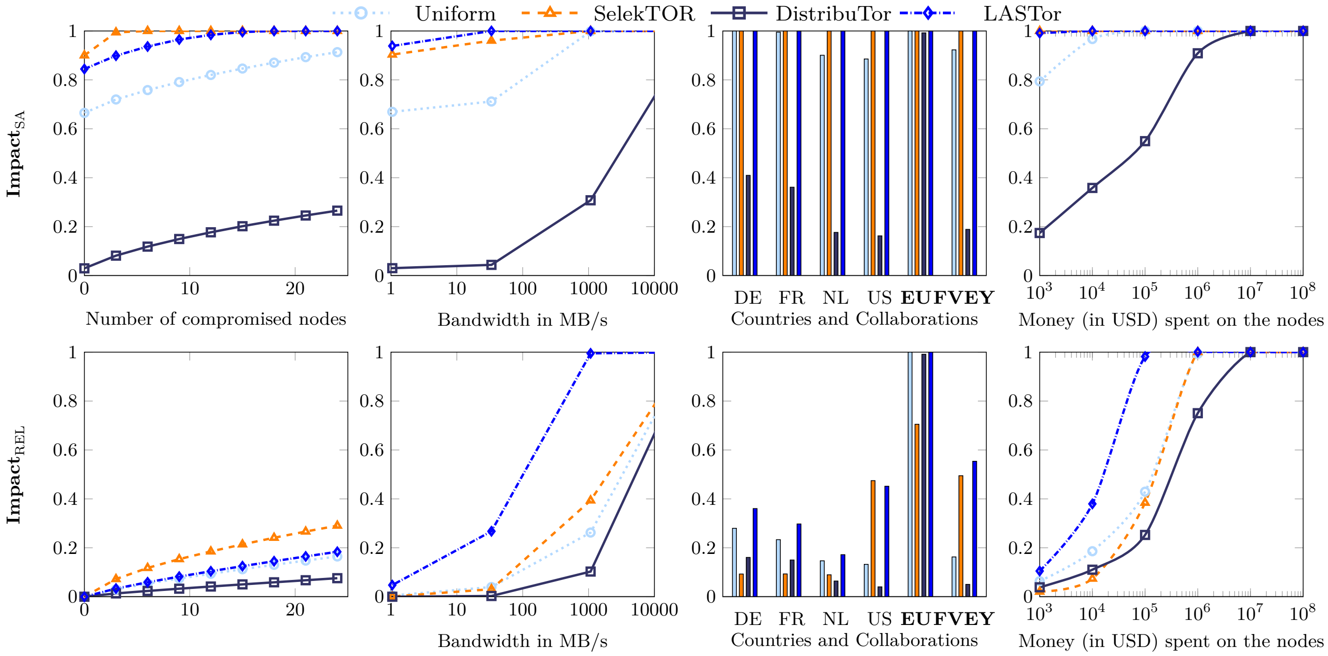


Fig. 10. $\textbf{Impact}_X$ for sender and relationship anonymity *during the transition phase* for different adversarial strategies (from left to right): k-collusion, bandwidth, geographic, and monetary.

References

[1] Ookla's NetIndex. http://www.netindex.com/value/allcountries/. Accessed July, 2015.

[2] Sourcecode of our analysis tool. https://www.infsec.cs.uni-saarland.de/projects/anonymity-guarantees/mator2.html.

[3] Tails - live operating system focused on privacy and anonymity. https://tails.boum.org/. Accessed February, 2015.

[4] The Tor blog, announcing the release of tor 0.2.6.10. https://blog.torproject.org/blog/tor-02610-released. Accessed August, 2015.

[5] Tor Metrics Portal. https://metrics.torproject.org/. Accessed July, 2015.

[6] Tor's Specification. https://gitweb.torproject.org/torspec.git/tree/tor-spec.txt. Accessed February, 2015.

[7] M. Akhoondi, C. Yu, and H. V. Madhyastha. LASTor: A low-latency AS-aware Tor client. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 476–490. IEEE, 2012.

[8] M. Backes, A. Kate, P. Manoharan, S. Meiser, and E. Mohammadi. AnoA: A Framework for Analyzing Anonymous Communication Protocols. In *Computer Security Foundations Symposium (CSF), 2013 IEEE 26th*, pages 163–178. IEEE, 2013.

[9] M. Backes, A. Kate, P. Manoharan, S. Meiser, and E. Mohammadi. AnoA: A Framework For Analyzing Anonymous Communication Protocols — Unified Definitions and Analyses of Anonymity Properties. IACR Cryptology ePrint Archive, Report 2014/087, 2014. available at http://eprint.iacr.org/2014/087.

[10] M. Backes, A. Kate, S. Meiser, and E. Mohammadi. (nothing else) MATor(s): Monitoring anonymity in Tors path selection algorithm. In *21st ACM Conference on Computer and Communications Security (CCS'14)*, CCS '14, pages 513–524. ACM, ACM, 2014.

[11] M. Backes and B. Köpf. Quantifying information flow in cryptographic systems. *Mathematical Structures in Computer Science*, 25(2):457–479, 2015.

[12] M. Backes, S. Meiser, and M. Slowik. Your choice mator(s): Large-scale quantitative anonymity assessment of tor path selection algorithms against structural attacks. Technical Report A/03/2015, Saarland University, December 2015.

[13] C. Diaz, S. Seys, J. Claessens, and B. Preneel. Towards measuring anonymity. In *Privacy Enhancing Technologies*, pages 54–68. Springer, 2003.

[14] R. Dingledine, N. Hopper, G. Kadianakis, and N. Mathewson. One fast guard for life (or 9 months). In *7th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2014)*, 2014.

[15] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. Technical report, DTIC Document, 2004.

[16] M. Edman and P. Syverson. As-awareness in Tor path selection. In *Proceedings of the 16th ACM conference on Computer and communications security*, pages 380–389. ACM, 2009.

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

[18] N. Gelernter and A. Herzberg. On the limits of provable anonymity. In *Proceedings of the 12th ACM workshop on*

Workshop on privacy in the electronic society, pages 225–236. ACM, 2013.

[19] A. D. Jaggard, A. Johnson, P. Syverson, and J. Feigenbaum. Representing network trust and using it to improve anonymous communication. In *In 7th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2014)*, 2014.

[20] A. Johnson, C. Wacek, R. Jansen, M. Sherr, and P. Syverson. Users get routed: Traffic correlation on Tor by realistic adversaries. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 337–348. ACM, 2013.

[21] S. J. Murdoch and R. N. M. Watson. Metrics for security and performance in low-latency anonymity networks. In N. Borisov and I. Goldberg, editors, *Proceedings of the Eighth International Symposium on Privacy Enhancing Technologies (PETS 2008)*, pages 115–132. Springer, July 2008.

[22] A. Neil. SelekTOR - Tor exit node selection made simple. http://www.dazzleships.net/?page_id=71. Accessed February, 2015.

[23] A. Serjantov and G. Danezis. Towards an information theoretic metric for anonymity. In *Privacy Enhancing Technologies*, pages 41–53. Springer, 2003.

[24] Y. Sun, A. Edmundson, L. Vanbever, O. Li, J. Rexford, M. Chiang, and P. Mittal. Raptor: routing attacks on privacy in tor. *arXiv preprint arXiv:1503.03940*, 2015.

[25] C. Wacek, H. Tan, K. S. Bauer, and M. Sherr. An empirical evaluation of relay selection in Tor. In *20th Annual Network & Distributed System Security Symposium (NDSS)*, 2013.

[26] T. Wang, K. Bauer, C. Forero, and I. Goldberg. Congestion-aware path selection for Tor. In *Financial Cryptography and Data Security*, pages 98–113. Springer, 2012.

A Appendix: Postponed Proofs and Figures

In this section we present the proofs for Theorems 1 and 2, and Lemma 1, as well as all postponed figures.

A.1 Proof of Theorem 1

Proof. To show the Theorem, we leverage the composition theorem of [10]. To this end, we use the original formulation of AnoA and consider budget adversaries as wrapper machines $A_f^B(\cdot)$ that internally run an arbitrary PPT adversary $\mathcal{A}$ and ensure that the corruption requests of $\mathcal{A}$ satisfies the constraints of a budget adversary for f and B. The composition requires that for every function f, every $B \in \mathbb{N}$ and for every anonymity function α, the adversary class A_f^B is composable as in Definition 3 from [10], i.e., that it satisfies the three properties *reliability* (the class does not start challenges on its own), *alpha-renaming* (the challenge counter holds no semantical meaning) and *simulatability* (the challenges are not handled structurally different than input messages).

Reliability: By construction, $A_f^B(\mathcal{A})$ sends messages $(\mathsf{challenge}, m)$ if and only if it receives a message $(\mathsf{challenge}, m)$ from $\mathcal{A}$. Thus, $A_f^B(\mathcal{A})$ is reliable.

Alpha-renaming: As the functional behavior of $A_f^B(\cdot)$ is completely agnostic to the challenge counter Ψ, $A_f^B(\mathcal{A})$ trivially satisfies alpha-renaming.

Simulatability: As, by construction, $A_f^B(\mathcal{A})$ only forwards challenge and input messages, we construct the following "trivial simulator" $\mathcal{S}^{\alpha}_{\mathrm{S}_0,\mathrm{S}_1,\mathrm{R}_0,\mathrm{R}_1}$ for any anonymity notion α and any two pairs of senders $\mathrm{S}_0, \mathrm{S}_1$ and recipients $\mathrm{R}_0, \mathrm{R}_1$. For a string $\vec{z} = [(z_1, b_1), \ldots, (z_n, b_n)] \in \{0,1\}^{2n}$, $\mathcal{S}^{\alpha}_{\vec{z}}$ behaves as follows. If $z_i = \mathsf{sim}$, it replaces all messages $(\mathsf{challenge}, m)$ by $(\mathsf{input}, \mathrm{S}^*, m, \mathrm{R}^*)$, where $(\mathrm{S}^*, \mathrm{R}^*) \leftarrow \alpha(\mathrm{S}_0, \mathrm{S}_1, \mathrm{R}_0, \mathrm{R}_1, b_i)$. For every $\vec{z} \in \{0,1\}^{2n}$, the simulator $\mathcal{S}^{\alpha}_{\vec{z}}$ satisfies the conditions from Definition 3 in [10].

Since $A_f^B(\mathcal{A})$ satisfies all three necessary conditions, it is composable. □

Remark: The proof for Theorem 1 also holds for the original, more complex description of the AnoA challenger from [8] in which the adversary chooses the senders and recipients for the challenges. In this case, the simulator computes α on the senders and recipients chosen by $\mathcal{A}$ instead of on the given fixed senders and recipients. Moreover, the proof is oblivious to the definition of the anonymity function α and also applies to the session definitions in [10].

A.2 Proof of Lemma 1

Proof. Let $\mathrm{S}_0, \mathrm{S}_1$ be two senders, $\mathrm{R}_0, \mathrm{R}_1$ two recipients, A as in the definition, and let $\{n\} \subseteq \mathcal{N} \cup \{\mathrm{S}_0, \mathrm{R}_0\}$ denote the output to $\textsc{Ch}^*$. Since $\textsc{Ch}^*$ only notifies the adversary if there is communication at n, any adversary in A hence only makes observations $o \in \mathit{Obs}$, i.e., $o \in \{\mathrm{S}_0, \mathrm{S}_1, \bot\} \times \mathcal{N}'^3 \times \{\mathrm{R}_0, \mathrm{R}_1, \bot\}$. Since $n \neq \mathrm{S}_0$ if $X \in \{\mathrm{SA}, \mathrm{REL}\}$ and $n \neq \mathrm{R}_0$ if $X \in \{\mathrm{RA}, \mathrm{REL}\}$, we have that the adversary makes observations at precisely one observation point n.

We only prove the lemma for sender anonymity; the adaptation to recipient anonymity and relationship anonymity is straightforward. We first divide the

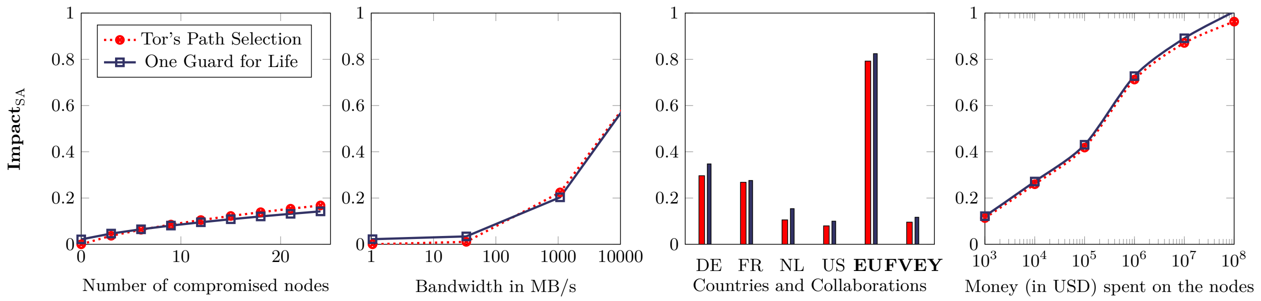


Fig. 11. Adversarial **Impact** against guard discovery attack, for different strategies (from left to right): k-collusion adversary, resource-constrained bandwidth adversary, geographic adversary, and monetary adversary.

set of all observations Obs into two subsets, depending on their probability and the challenge bit. We set $Obs_0 := \{o \in Obs | \Pr[o = \mathcal{O}[\{n\}](c), c \leftarrow \mathsf{ps}(\mathrm{S}_0,\mathrm{R}_0)] > \Pr[o = \mathcal{O}[\{n\}](c), c \leftarrow \mathsf{ps}(\mathrm{S}_1,\mathrm{R}_0)]\}$, and $Obs_1 := Obs \setminus Obs_0$. An adversary $\mathcal{A}$ hence maximizes its advantage by outputting 0 if and only if it makes an observation $o \in Obs_0$. For $i \in \{0,1\}$, we obtain

$$\begin{aligned}&\Pr[0 \leftarrow \langle \mathcal{A}(1^\eta)||\textsc{Ch}(\alpha,1,\mathrm{S}_0,\mathrm{S}_1,\mathrm{R}_0,\mathrm{R}_1,i)\rangle]\\ =&\sum_{o\in Obs_i} \Pr[o = \mathcal{O}[\{n\}](c)|c \leftarrow \mathsf{ps}(\mathrm{S}_i,\mathrm{R}_0)]\end{aligned}$$

We are left to show that the advantage of this adversary $\mathcal{A}$ is equal to $\mathbf{Impact}^{\mathsf{obs}}_{\mathrm{SA}}(n)$:

$$\begin{aligned}&\Pr[0 \leftarrow \langle \mathcal{A}||\textsc{Ch}(\alpha,1,\mathrm{S}_0,\mathrm{S}_1,\mathrm{R}_0,\mathrm{R}_1,0)\rangle]\\ &-\Pr[0 \leftarrow \langle \mathcal{A}||\textsc{Ch}(\alpha,1,\mathrm{S}_0,\mathrm{S}_1,\mathrm{R}_0,\mathrm{R}_1,1)\rangle]\\ &=\sum_{o\in Obs_0} \Pr[o = \mathcal{O}[\{n\}](c)|c \leftarrow \mathsf{ps}(\mathrm{S}_0,\mathrm{R}_0)]\\ &-\sum_{o\in Obs_1} \Pr[o = \mathcal{O}[\{n\}](c)|c \leftarrow \mathsf{ps}(\mathrm{S}_1,\mathrm{R}_0)]\end{aligned}$$

$$\begin{aligned}=&\sum_{o\in Obs} \phi\Big(\Pr[o = \mathcal{O}[\{n\}](c)|c \leftarrow \mathsf{ps}(\mathrm{S}_0,\mathrm{R}_0)],\\ &\qquad\Pr[o = \mathcal{O}[\{n\}](c)|c \leftarrow \mathsf{ps}(\mathrm{S}_1,\mathrm{R}_0)]\Big)\\ =&\,\mathbf{Impact}^{\mathsf{obs}}_{\mathrm{SA}}(n),\end{aligned}$$

where $\phi(Y,Z)$ is defined as $Y-Z$ if $Y > Z$, and 0 otherwise. □

A.3 Proof Sketch for Proof of Theorem 2

The proof of Theorem 2 follows the intuition of the proofs in [10], but requires less approximation for the individual impacts of the nodes. By Theorem 1 and Lemma 22 from the full version of the AnoA framework [9], it suffices to show that the ideal functionality of Tor, $\mathcal{F}_{\mathrm{OR}}{}'$, is $(\delta,1)$-IND-ANO for $\delta = \mathbf{Impact}_X(A^B_f)$. Once we showed this, we immediately obtain that Tor is $(\delta,1)$-IND-ANO for $\delta = \mathbf{Impact}_X(A^B_f)$ plus a negligible additive factor, since Tor constitutes a UC-secure realization of $\mathcal{F}_{\mathrm{OR}}{}'$. Roughly, $\mathcal{F}_{\mathrm{OR}}{}'$ does not send actual onion encryptions, but only provides handles over the network and thereby eliminates cryptographic objects from the construction. For compromised nodes, $\mathcal{F}_{\mathrm{OR}}{}'$ reveals which of these handles belong together. By Lemma 1 we know that any observation made by any individual observation point n impacts anonymity by exactly $\mathbf{Impact}^{\mathsf{obs}}_X(n)$ for the anonymity notion α_X under consideration. Intuitively, the proof divides the set of all observations into distinct subsets of observations, depending on where compromised nodes sit in a circuit. Then, for every such set, we compare the impact of each observation if a set of Tor nodes (and the malicious recipient/sender) is compromised with the sum of the impacts of all compromised Tor nodes (and the malicious sender/recipient) on their own. Since we sum over all these Tor entities, for the majority of observations the impact of the sum is larger than the impact of the combined set. However, the lack of observation of certain (compromised) nodes can increase the impact that other compromised nodes have on other compromised Tor entities. We calculate this indirect impact and yield the formulas from Figures 3 to 5.

We refer the reader to the extended version of this paper [12], as the full proof is too long to fit into this version.

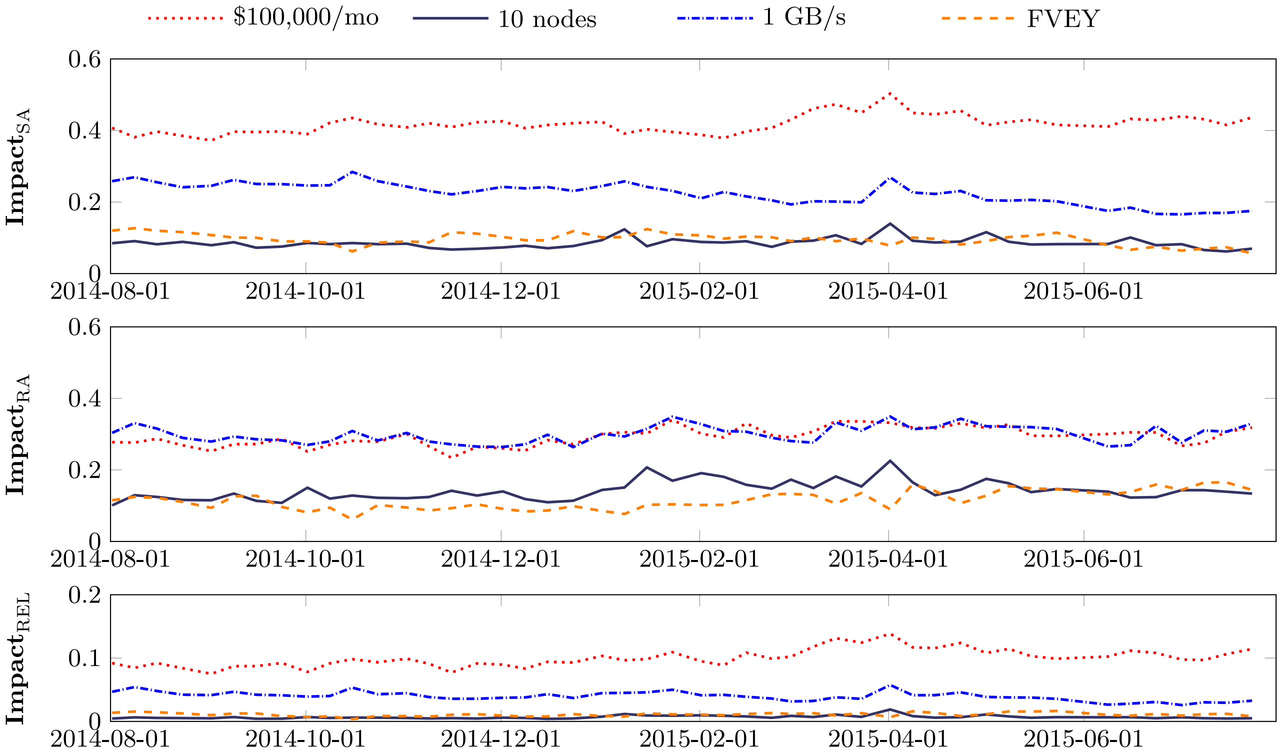
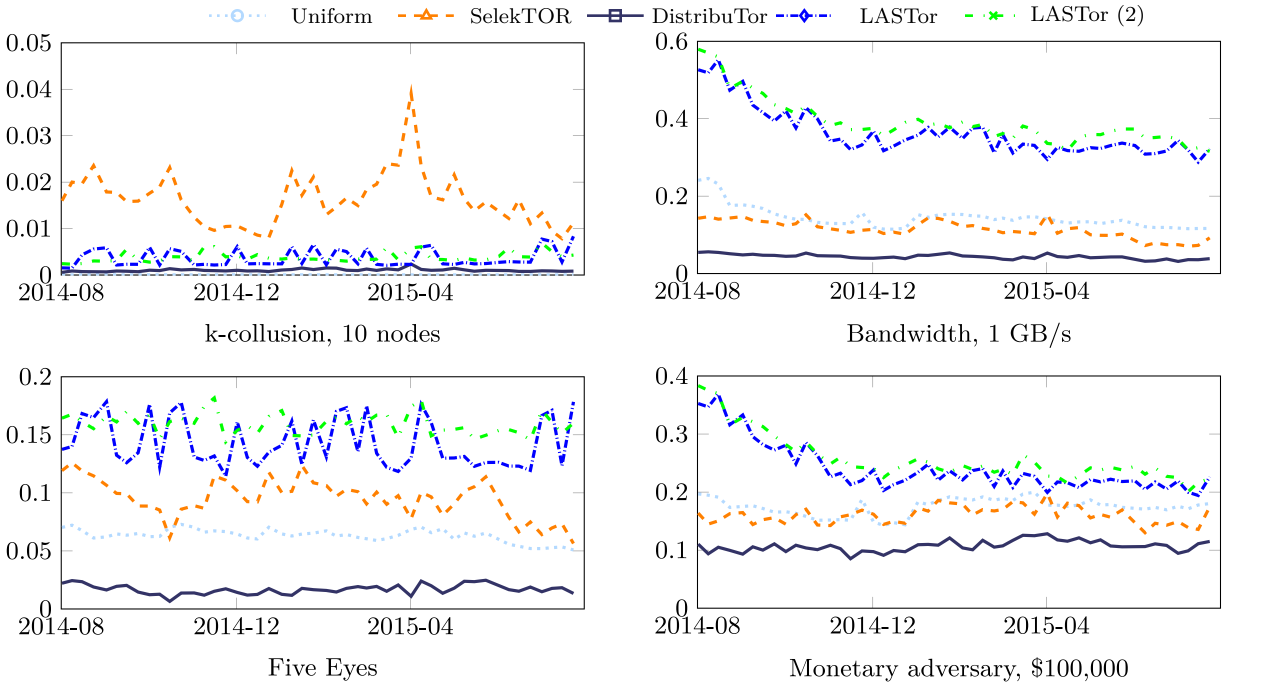


Fig. 12. Changes in the adversarial **Impact** against selected adversarial strategies for Tor's path selection during the last year. Note, that the Y axis scales from 0 to 0.6 for sender and recipient anonymity plots, and from 0 to 0.3 for relationship anonymity.



k-collusion, 10 nodes

Bandwidth, 1 GB/s

Five Eyes

Monetary adversary, $100,000

Fig. 13. Changes in the adversarial **Impact** against relationship anonymity for different strategies (from top left): k-collusion adversary compromising 10 nodes, resource-constrained bandwidth adversary compromising 1 GB/s, Five Eyes countries and monetary adversary with budget of $100,000 per month, for alternative path selection algorithms. **Note**, that each plot has an individual scale on Y axis to improve readability.

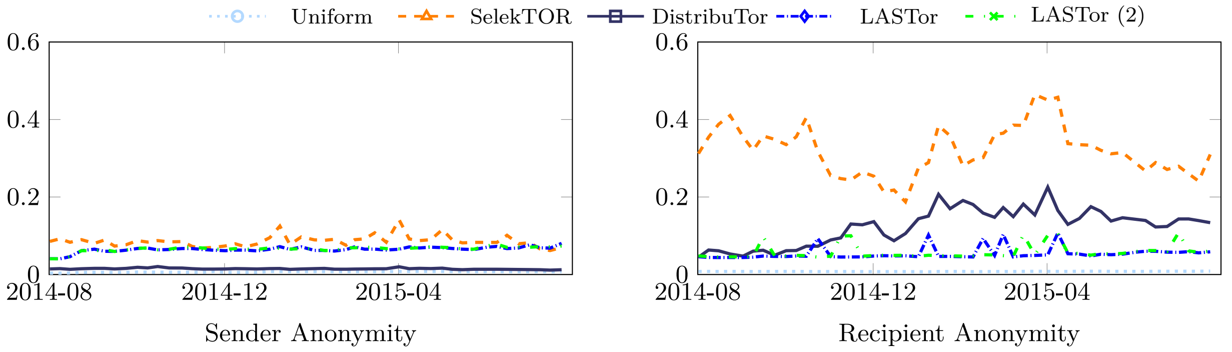


Fig. 14. Changes in the adversarial **Impact** of k-collusion adversary compromising 10 nodes, in sender and recipient anonymity, for alternative path selection algorithms.

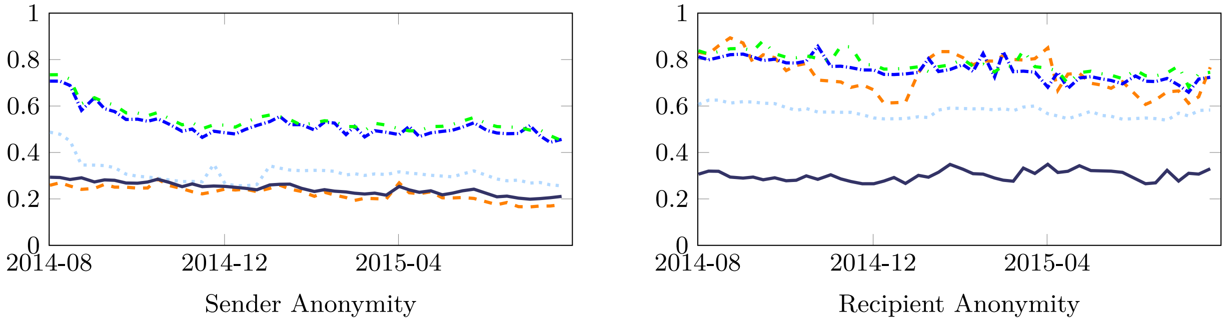


Fig. 15. Changes in the adversarial **Impact** of resource-constrained bandwidth adversary compromising nodes of total bandwidth up to 1 GB/s, in sender and recipient anonymity, for alternative path selection algorithms.

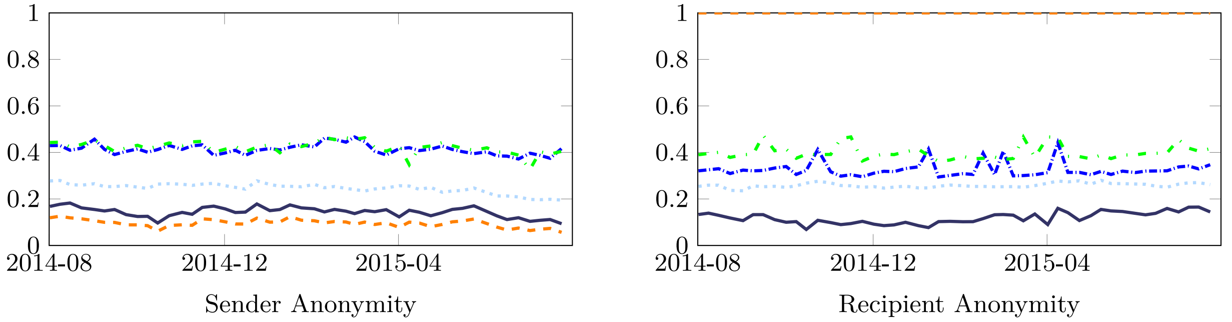


Fig. 16. Changes in the adversarial **Impact** of Five Eyes countries, in sender and recipient anonymity, for alternative path selection algorithms.

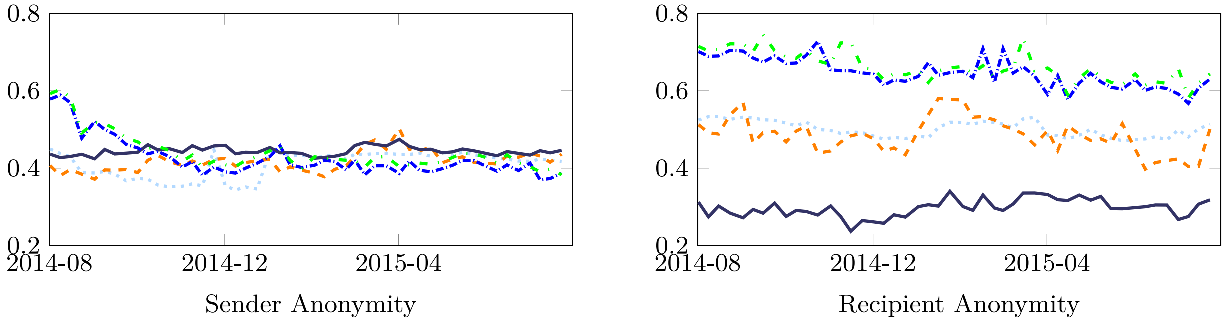


Fig. 17. Changes in the adversarial **Impact** of monetary adversary with budget $100,000, in sender and recipient anonymity, for alternative path selection algorithms.