

Quantifying Resistance to the Sybil Attack

N. Boris Margolin and Brian Neil Levine

Dept. of Computer Science, Univ. of Massachusetts, Amherst, MA USA
{margolin, brian}@cs.umass.edu

Abstract. Sybil attacks have been shown to be unpreventable except under the protection of a vigilant central authority. We use an economic analysis to show quantitatively that some applications and protocols are more robust against the attack than others. In our approach, for each distributed application and an attacker objective, there is a critical value that determines the cost-effectiveness of the attack. A Sybil attack is worthwhile only when the critical value is exceeded by the ratio of the value of the attacker’s goal to the cost of identities. We show that for many applications, successful Sybil attacks may be expensive even when the Sybil attack cannot be prevented. Specifically, we propose the use of a recurring fee as a deterrent against the Sybil attack. As a detailed example, we look at four variations of the Sybil attack against a recurring fee based onion routing anonymous routing network and quantify its vulnerability.

1 Introduction

Many distributed services and peer-to-peer (p2p) applications are vulnerable to *Sybil attacks* [17], where a single malicious *entity* masquerades as many counterfeit *identities* and uses them to launch a coordinated assault. The attack can be used to ruin the integrity of reputation systems [5, 12], create false routes in mobile ad hoc networks [23], identify users of anonymous routing protocols [16], cheat p2p computing systems (e.g., SETI@home) [47], and free-ride cooperative file storage systems [15]. A form of the Sybil attack is commonly used to fool Google’s PageRank algorithm [6]. In most situations it is not possible to prevent Sybil attacks using resource tests, and certificate systems generally do not guarantee no entity has two keys. The Sybil attack has been widely studied but remains unsolved in general. Several papers have evaluated formally the conditions under which applications are susceptible to the attack [12, 17], however this is a coarse-grained approach. It is not true that all applications are equally vulnerable.

In this paper, we quantify the threat of the Sybil attack using an economic model. For the first time, we show that the attack poses a different level of threat to different applications. We derive a concrete measure of attack resistance called the *Sybil valuation*. The valuation is the critical ratio of the value of the attacker’s goal to the per-identity cost of the protocol — at the critical ratio or above the attacker can expect to profit from attacking the protocol. This measure allows us to quantitatively compare the threat as

This work was supported in part by National Science Foundation award NSF-0133055.

the number of peers in the system changes. Moreover, our analysis distinguishes applications not by what type of service they provide, but rather by the specific Sybil attack variations the applications allow.

Many have suggested addressing the Sybil attack by charging one-time resource costs per participating identity. As part of our analysis, we show that protocols that charge a *recurring fee per participating identity* are more effective as a disincentive against successful Sybil attacks. Our previous work on the Informant protocols is an example [26] of how recurring fees can also be used to detect Sybil attacks, and our work here is complementary. We show that recurring fee protocols are more secure because they require that the successful attacker's has resources that scale linearly with the number of other participants (instead of a constant amount). This linear requirement is synergistic with p2p applications that seek to increase the number of peers for other performance benefits. Moreover, in recurring fee protocols, two uncoordinated Sybil attackers will increase the resources required of each other without increasing costs to honest participants.

As a concrete example of protocol analysis using the Sybil valuation, we evaluate a Recurring Fee Onion Routing protocol, which we refer to as *RFOR* to distinguish it from Onion Routing or the deployed Tor protocol [16], which operates without explicit or implicit recurring fees. In Tor, as it is deployed now, an attacker will always find benefit from setting up just two identities no matter the population size, and the attacker can amortize all costs over time to a negligible amount. Evaluations of RFOR using the traces of participation of the actual Tor system show a sharp contrast. For example, as of September 2007, the traces show $n = 1373$ volunteer peers acting as proxy servers. With the recurring router entry fee charged by a RFOR protocol, a rational entity would have to value the knowledge of a single connection of a specific user at $4n$ times the router entry fee in order to launch a Sybil attack. For a fee set at \$0.01, this value is \$54.92 given the population in the traces, which may be enough to discourage only casual attackers; if RFOR was deployed to protect users that are more concerned about their anonymity, the fee could be set higher. Using these real traces in our evaluation, we are able to show that a RFOR system would grow less vulnerable to some Sybil attacks with increases popularity, but would still be susceptible to Sybil-based DoS attacks by resource-poor attackers.

For applications that cannot tolerate any entity with multiple identities, centralized manual identity certification is the only solution, but, as Douceur points out, few applications can bear the cost. For many applications, managing recurring payment of fees is a more reasonable solution. Several distributed, Internet-based micro-payment schemes can manage fees [7, 8, 41, 42]. For applications whose users may be unwilling to make monetary payments, fees can also be imposed less robustly, though perhaps more readily, through the use of non-monetary mechanisms such as CAPTCHAs [43] and SMS messages. Our approach is flexible in that peers in the system need to show only that a payment was made; the payment does not have to be to other peers in the system. Furthermore, in our approach identities are never asked to prove they are separate entities.

Outline. In Section 2, we state our model and assumptions regarding identity, protocols, and Sybils. In Section 3, we present a cost-benefit analysis for malicious attackers based on entry fees. In Section 4 we discuss different types of entry fees. In Section 5, we give

an overview of approaches to the Sybil attack in the literature. To our knowledge, this is the first broad overview of research on this subject. We offer concluding remarks in Section 6.

2 Goals and Model

Our model is an extension of Douceur’s [17] in which each peer participating in a network protocol is a unique *identity* that is controlled by a rational actor [34] known as an *entity*. A Sybil attack occurs when one entity secretly controls multiple identities. Identities send messages to each other through a *communications cloud* that precludes definite identification using direct observation. We assume that messages can be securely linked to identities, though not to entities. This can be accomplished, as in Douceur, by having identities choose public/private key pairs and signing their messages or by other methods [37]. The use of public keys does not imply a PKI because the keys are not linked securely to any real-world entity.

We model the applications running the network protocol as having an *entry phase* and a *service phase*. During the entry phase, each identity is charged an *entry fee*, and we assume the identity can later demonstrate to others that the fee has been paid. Recurring fee applications force peers to repeat the entry phase (and fee) after one or more service phases.

Below, we introduce a model that includes entity utility and strategy, the value of the Sybil attack in general, and then we define three specific types of Sybil attacks. First, we discuss the limitations of our approach.

Limitations. Our model applies to applications that involve weakly authenticated participants sharing resources. We show that such applications that charge recurring fees are more secure than those that pay a one-time fixed cost. We evaluate anonymous communication systems (and other applications) below as an example — yet, Tor charges no fees at all currently. It is not our intent to compare having no fees (and therefore no defense against the Sybil attack) against having fees. Moreover, our analysis does not indicate whether applications would be more or less popular if they charged (recurring) fees. On the one hand, some users might not find the increased cost worth the application’s services; on the other hand, some users might find the application has added benefit since it is more secure. The answer to this financial question depends on the specific application and business model.

We do not investigate how to ensure fees are paid, though many others have done so [7,8,41,42]). We do note that doing so is an easier task than requiring a trusted authority that can certify that each identity is an independent entity; the latter is difficult even with access to real-world documents [1]. Finally, we note that fees do not need to be monetary, and typically will not be. Instead, it may involve the use of CAPTCHAS [43], SMS messages, or other techniques, as discussed in Section 4. Given the prevalence of botnets, fees that can be paid by obtaining a computer and IP address are not satisfactory.

2.1 Entity Utility

Because our entities are rational actors, they have a specific utility for each possible protocol outcome, and they apply strategies that give them the highest possible expected utility. Rational actors perform a cost-benefit analysis to determine what action to take — including whether to launch a Sybil attack against a specific protocol.

Our model follows basic game theory [34]. Let E represent a set of entities participating in a protocol, controlling a set of identities I . Let S_e be the set of possible actions, called *strategies*, that an entity $e \in E$ can carry out. An entity must decide on a single strategy based on his knowledge and goals. An example strategy would be launching a Sybil attack with a certain number of identities. Since there are multiple entities participating, there is a set of outcomes for $n = |E|$ entities

$$O = S_{e_1} \times S_{e_2} \times \cdots \times S_{e_n} \quad (1)$$

The combination of the strategies of participating entities completely defines an *outcome*. An outcome $o \in O$ is a selection of one strategy from each of these n sets; that is o is tuple $(s_{e_1}, \dots, s_{e_n})$ representing the strategy taken by all entities. For simplicity, we do not discuss non-deterministic (i.e., irrational) attackers, but they require only minor changes to our model.

Each entity's preferences are expressed using a utility function that maps outcomes to a utility score. The utility of an outcome o to an entity e is the sum of a *benefit utility* $\tau_e(o)$ and a *cost utility* $\pi_e(o)$ (normally negative) determined by payments made by e in outcome o :

$$u_e(o) \equiv \tau_e(o) + \pi_e(o). \quad (2)$$

When entry fees are used, the cost, $\pi_e(o)$, is the product of the entry fee and the number of identities controlled by the entity.

2.2 The General Sybil Objective

For an attacker entity m considering the wisdom of a Sybil attack, $\sigma_q \in S_m$ represents the strategy of entering q identities — and doing whatever else is necessary in order to reach some objective.

Let A be the set of the objectives that an attacker can attempt to achieve using Sybil attacks. We define an *objective success count* operator $\psi(o)$, which gives the number of successes by m in the outcome o . For example, one set of objectives is to control the entire path through an anonymity system, revealing the initiator of a packet. When participating as multiple Sybil identities, an attacker may control multiple paths, revealing multiple initiators, which increases the value of ψ accordingly.

We assume that the attacking entity $m \in E$ values attacks linearly, with the success of a single attack valued at v , so that

$$\tau_m(o) = v\psi(o). \quad (3)$$

In general, an attacker's expected benefit from a Sybil attack using q identities is

$$\mathbb{E}[\tau_m | s_m = \sigma_q] = \sum_{o \in O} v\psi(o) \Pr[o | \sigma_q]. \quad (4)$$

We restrict our analysis to protocols in which honest entities do not gain any benefit from Sybil attacks. That is, we assume honest users value most the protocol’s objectives (e.g., anonymity) and that malicious users value outside objectives more (e.g., breaking anonymity).

2.3 Specific Sybil Objectives

The Sybil attack can be launched as one of several specific *objectives* that depend on the application being attacked. We distinguish these attacks by the way the application uses peers to offer service. In all cases, the application starts the service phase by selecting a subset of k peers (identities); typically, a subset k is selected for each of the n participating identities. For example, Crowds forms a path of k peers for anonymous routing for each peer that is a source of traffic. From here, we can distinguish several different Sybil attack objectives.

First, for any specific application, there is a **minimum number of identities** required for a successful attack. For example, to successfully launch the predecessor attack an attacker needs only $c = 1$ identities for the Crowds protocol but $c = 2$ identities to for the Onion Routing protocol [45].

Second, we distinguish **One-time fee objectives**, which are applicable to applications where the attacker can launch Sybil attacks repeatedly without additional cost — as when entry fees are charged only one time ever per identity. Since any attack with a non-zero probability of success is expected to succeed eventually, a given strategy has either no chance of success, or is guaranteed success. In this case, the only strategies that the attacker needs to consider are σ_0 (entering no identities) and σ_c (entering the minimum number of identities required for success). One-time fee attacks are denoted T_c . Onion Routing and the deployed Tor system are examples of one-time fee protocols.

Third, for applications that charge a **recurring entry fee** for one or more service phases, several attacks can be distinguished. For example, while most anonymous routing protocols create subgroups (paths) of k peers, choosing with replacement from a set I of peers, Pastiche [15] is a p2p application that stores backup data from each source node with k other peers, choosing without replacement (though Pastiche does not charge fees in reality). In both cases, the objective of the attacker is to control c of the k identities chosen each service phase for each of n sources. However, we distinguish the former case as a *binomial objective*, since k identities are chosen *with* replacement from I . And we refer to the latter case as a *hypergeometric objective*, since k identities are chosen *without* replacement from I . We further detail these cases below.

- **Binomial objectives.** For each identity in the application, a subgroup is chosen with replacement, and the attacker may try to target all subgroups, a specific victim’s subgroup, or try to succeed against any (that is, no one specifically) victim’s subgroup, as we detail below.
 - When attacking a **specific** subgroup, denoted $B_{c,k}^{\text{spec}}$, the attacker’s utility is proportional to the probability of success against the one identity.
 - When seeking success against **any one** subgroup, denoted $B_{c,k}^{\text{any}}$, the attacker’s utility is proportional to the probability of success against at least one identity.

- When attacking **all** n subgroups, denoted $B_{c,k}^{\text{all}}$, the attacker’s utility is proportional to the total number of group control successes.
- **Hypergeometric objectives.** A subgroup of k identities are chosen without replacement. Such objectives are denoted $H_{c,k}$. SETI@home [38] and Pastiche [15] are subject to the hypergeometric objectives, since identities in peer groups are chosen without replacement, for redundancy. The notation $H_{c,k}$ represents the objective of controlling a specific peer group. There are a large number of natural subcases of the Hypergeometric objective (compared to just three for the Binomial objective), and to avoid complexity we omit them.

Commonly, p2p applications select identities for subgroups uniformly at random, and we assume so here for all objectives; our previous work [45] suggests that uniform random selection is the most attack-resistant approach for anonymous communications systems, and we conjecture that it is the most attack-resistant approach for many other p2p systems as well.

3 The Sybil Valuation

In this section, we use our model to determine when the benefits of a specific Sybil attack exceeds the costs, a point we call the *Sybil valuation*. *When an attacker’s valuation of their objective, in terms of the entry fee cost, exceeds the Sybil valuation, it is in their interest to launch the attack.*

We denote the Sybil valuation for an objective a by γ_a , defined

$$\gamma_a \equiv \min_q \frac{q}{E[\psi_a|\sigma_q]} \quad (5)$$

where $E[\psi_a|\sigma_q]$ gives the expected number of successes for an attacker with the objective a launching a Sybil attack with q identities.

Using this measure, a protocol designer or user can determine how intrinsically resistant a protocol is to a Sybil attack, so she can independently evaluate the design of the protocol and the setting of entry fee. Once the design is fixed, she can use the measure to determine how to set the entry fee to discourage attackers with different valuations for success in reaching an objective.

First, we show that an attacker m only benefits from an attack when their objective valuation is at least γ_a times their per-identity cost. The attacker’s expected utility for a Sybil attack with q identities must be non-negative for the attack to be rational. So an attack is rational if and only if

$$E[\tau_m|\sigma_q] - qf \geq 0 \quad (6)$$

$$vE[\psi_a|\sigma_q] \geq qf \quad (7)$$

$$v \geq \frac{qf}{E[\psi_a|\sigma_q]}. \quad (8)$$

Objective Type	Example Applications	Specific objective	Optimal num identities	Sybil Valuation (γ_a^*)	γ_a^* as $n \rightarrow \infty$
One-time Fee	[13, 17, 24]	T_c	c	c	c
Binomial	RFOR denial of service, RFOR endpoints attack, and Predecessor attack [44]	$B_{1,k}^{\text{spec.}}$	1	$(1 - (\frac{n}{n+1})^k)^{-1}$	$k^{-1}n$
		$B_{k,k}^{\text{spec.}}$	$(k-1)n$	$(\frac{k}{k-1})^{k-1} kn$	ekn
		$B_{1,k}^{\text{any}}$	1	$(1 - (\frac{n}{n+1})^{kn})^{-1}$	e^{-k}
		$B_{k,k}^{\text{any}}$	$(k-1)n + k$		kn
		$B_{1,k}^{\text{all}}$	1	$1/n(1 - (\frac{n}{n+1})^k)$	k^{-1}
Hyper-geometric	SETI@Home [38], Pastiche [15]	$H_{1,k}$	Same as $B_{1,k}^{\text{spec.}}$		
		$H_{k,k}$	$(k-1)n + k$		kn

Table 1. Optimal number of identities, Sybil valuation, and asymptotic behavior as n grows large for different objectives. For derivations, see the Appendix. The Sybil valuations of the Binomial and Hypergeometric objectives have no closed-form representation for general c and k [11], and they are omitted for readability.

Since the attacker is rational, she will choose the optimal number of identities q to include in the protocol. Therefore,

$$v \geq \min_q \frac{qf}{\mathbb{E}[\psi_a | \sigma_q]} \quad (9)$$

$$v \geq \gamma_a f. \quad (10)$$

For clarity, we began this subsection by defining the Sybil valuation; note that Inequalities 6 through 9 are a template for deriving the ratio.

Inequality 10 says nothing about the *resources* available to a particular attacker. An attacker may value an objective highly, but not launch a Sybil attack if she does not have enough sufficient resources to achieve it. However, in this paper, we take the defender’s point of view and conservatively assume that an attacker controls an unlimited amount of resources.

In some cases, the optimal number of identities q will be very small, so the attacker will only have a very small chance of success each round. By entering a larger number of identities, the attacker would decrease the expected number of rounds until success, but the expected total cost would be higher.

3.1 γ_a for Specific Objectives

We now quantify γ_a , the susceptibility of applications to Sybil attacks. Table 1 has results for each objective discussed in Section 2.3: the one-time fee, binomial, and hy-

pergeometric objectives. While the derivation of γ_a is not difficult, the analysis of each protocol type is more involved, lengthy, and in some cases it has no closed form.

As an example, consider the objective $B_{1,k}^{\text{spec}}$, where a specific identity's subgroup is targeted and the attacker needs to be selected as only 1 of k peers in the subgroup. The probability of success given q identities is $1 - (\frac{n}{q+n})^k$. Therefore for this objective,

$$\gamma_a = \min_q \frac{q}{1 - (\frac{n}{q+n})^k}. \quad (11)$$

The minimizing q must be either 1 (the lowest possible value for q) or some root of the derivative of the minimized expression. It is possible to show that the derivative of the minimized expression is always positive for positive integer values q . The minimum must therefore be at 1, the lowest possible value for q , and therefore for this objective

$$\gamma_a = \frac{1}{1 - (\frac{n}{n+1})^k}. \quad (12)$$

As n grows large in this case, then γ_a approaches n/k , meaning that increasing popularity increases costs linearly for the attacker. Because of space limitations, details of the other γ_a calculations appear in our technical report [27]. We discuss the implications of the results that are summarized in Table 1 below.

- **The one-time fee objective**, T_c is easily achieved in most cases; regardless of the number of participants, it takes only c times the entry fee to achieve the objective. For example, the analysis applies to an onion-routing system requiring a one-time entry fee where the objective is the predecessor attack [44], which requires a minimum of two identities for success. Then an attacker only needs to value the attack at twice the entry fee and enter two identities into the protocol, which is a very inexpensive Sybil attack. One-time fees are not well-suited to discouraging Sybil attacks.

- **The binomial objective** varies in difficulty depending on the objective; the intended victim can be some specific user, any user, or all users.

Against specific users, the difficulty of achieving the binomial objective is linear in the n : a protocol is increasingly secure as more identities participate. This is true regardless of c and k , though c determines if γ_a is linear in k , linear in $1/k$, or somewhere in between¹.

In binomial objectives where the attacker wishes to succeed against any single user, c determines the difficulty of the attack. For $c = 1$, we find that γ_a converges to $\frac{e^k}{e^k - 1}$ as n increases. Therefore, in this case, adding more honest identities has limited benefit. Conversely, when $c = k$ (and $k > 1$), we find γ_a asymptotically approaches $(k - 1)n$ as n increases.

In binomial objectives including all users, γ_a is asymptotically constant with increasing n . For $c = 1$ it approaches $1/k$, while for $c = k$ it does not depend on n at all, but is asymptotically equal to ek .

- **Hypergeometric objectives** are those where an attacker attempts to control c of k peer group identities, chosen without replacement. They are similar to binomial

¹There is no closed-form expression of γ_a for any binomial objective when c is not exactly 1 or k ; see Casella and Berger [11].

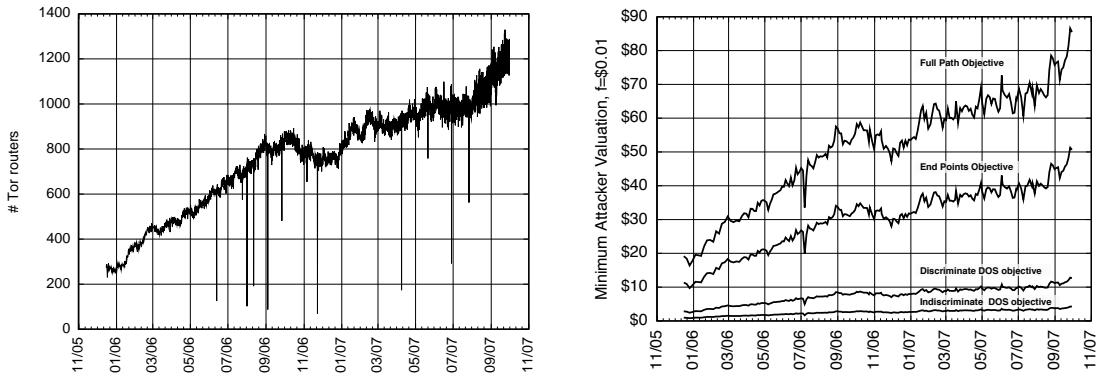


Fig. 1. (a) Tor router population over time. (b) Minimum valuation of four attacks against RFOR when $f = \$0.01$ (Population dips are smoothed out.)

objectives, but are more difficult for the attacker, since her identities cannot be reused; the difference is most pronounced when n and k are small.

3.2 Application: Recurring Fee Onion Routing (RFOR)

In this section, we apply our Sybil valuation measure to reveal properties of a recurring fee onion routing protocol. RFOR operates exactly according to the Onion Routing protocol definition with two exceptions; (i) routers pay a fee for every path reformation; (ii) paths are constructed by choosing proxies with replacement.

Our goal is to contrast RFOR's relative strength against the Sybil attack with standard Onion Routing, which provides free services to users through volunteer routers. There is no defense against the Sybil attack in Onion Routing in design or in various deployments; i.e., Sybil attacks can launch the attack successfully for a negligible cost.

Tor is an example of Onion Routing that is deployed with other defenses such as guard nodes. We note that Tor operators do pay a cost to operate a Tor node, but this cost is not one that would grow if those operators increased the number of Sybil identities they operate on the network from the same computer. For example, Murdoch discovered through clock skew analysis that 30 particular routers on the Tor network were actually just two real machines [30].

Our goal is not to ask if Onion Routing deployments such as Tor should charge users a fee, but rather *what is the cost that Sybil attackers should expect to pay in a recurring fee version of Onion Routing?* We are unaware of previous work that quantifies the threat posed by Sybil attacks (rather than collaborating entities) to Onion Routing.

As a simple example, we assume that RFOR routers are charged a fee of $f = \$0.01$ every path reformation. We assume that the *path* length (i.e., circuit length) remains at the default setting of three. Below, we analyze four objectives of RFOR Sybil attackers: two types of denial-of-service (DoS) attacks, capturing the endpoints of a path, and capturing the full path. We show that these objectives vary considerably in difficulty;

some have asymptotically linear Sybil valuations with reasonable coefficients, while others have asymptotically constant, and sometimes very low, Sybil valuations.

In our evaluations, we use Tor’s directory server’s public advertisements of available proxies, which have been archived by Peter Palfrader, who generously shared the data with us. The 73,309 trace files cover December 2005 until September 2007. See Wright et al. [45] for characteristics of Tor measurements, including up- and down-time distributions. Note that each peer router supports services to hundreds of clients of Tor. To join the Tor network as a routing peer, a person needs only a computer and one IP address for each router they wish to control. We are specifically concerned with attacking the peer routers that service the clients.

Figure 1(a) shows the number of Tor routers over time from the traces and Figure 1(b), discussed below, shows the required valuations of the objectives over time when there is a \$0.01 per path reformation fee for routers in RFOR. In our examples, we make use of peak population value in the logs of $n = 1,373$ routers on September 23, 2007. However, it is instructive to examine these values in Figure 1(b) in January 2006, when $n = 275$, approximately.

- **Discriminating DoS objective.** If an attacker can control a single router out of the three in a user’s RFOR path, she can deny service. In the discriminating DoS objective, the attacker wants to limit DoS to cases when there is a reasonable probability that the targeted user is on the target path. Specifically, the attacker only launches the DoS if she observes the target node as the previous node in the path. In this case there is a $1/3$ chance that the path was initiated by the target user. (If the attacker needs certainty, then the end-points objective, discussed below, applies.)

This objective corresponds to $B_{1,1}^{\text{spec.}}$. γ_a for the objective is $n + 1 = 1374$. So the attacker would need to value the attack at \$13.74 in order to decide to launch it.

- **Indiscriminate DoS objective.** In this case, the attacker launches a DoS attack even if the target is not observed as a predecessor, possibly causing collateral damage. This is the $B_{1,3}^{\text{spec.}}$ objective when she receives utility from victimizing only a specific user, and it is the $B_{1,3}^{\text{any}}$ and $B_{1,3}^{\text{all}}$ objectives otherwise.

For the $B_{1,3}^{\text{spec.}}$ objective, when the attacker has a specific targeted user, γ_a is $\frac{1}{1 - (\frac{n}{n+1})^3}$. At 1373 routers $\gamma_a = 458$, requiring a valuation of \$4.58 for the attack at a \$0.01 per-identity fee.

The cost is even less for the other objectives. The objective when the attacker is content with denying service to any one user, $B_{1,3}^{\text{any}}$, has a $\gamma_a = 1.05$ at 370 identities, requiring a valuation of \$0.0105 at a \$0.01 per-identity fee. The objective when the attacker receives utility that is proportional to the total number of users it can deny service to, $B_{1,3}^{\text{all}}$, has a $\gamma_a = 0.334$, requiring a valuation of just \$0.0033 for a profitable attack when the per-identity fee is \$0.01.

- **Endpoints objective.** For this objective, the attacker uses its sybil identities to capture the two end points of a path. The two proxies then launch a timing attack [31] to determine if they are on the same path, thereby learning the initiator and responder. The endpoints objective is a $B_{2,2}$ objective.

When the fee is charged per path reformation, the results are as follows. For the binomial objective, when the attack profits only from a specific user $\gamma_a = 4n$, or 5492 at $n = 1373$, which gives an attacker Sybil valuation of \$54.92. When the attacker

succeeds after revealing any one user as the initiator, we have $\gamma_a = \frac{n+2}{1-(1-(1-\frac{n}{2(n-1)})^2)^n}$, or about 1375 at $n = 1373$; so the Sybil valuation is \$13.75. When the attack profits from attacking all users, we have $\gamma_a = 4$. In this case the attacker only needs to value the objective at \$0.04.

• **Full-path objective.** Attackers attempt, in this case, to control all k nodes in the path, and can then know for certain that the endpoints are communicating without additional mechanisms. (We note that since RFOR makes no attempt to thwart timing attacks, and a more accurate analysis of RFOR’s vulnerability when using fees is given by the endpoints objective.)

We first analyze the full-path objective considering a specific user. If the fee of \$0.01 is charged per path reformation, then the objective corresponds to $B_{3,3}^{\text{spec}}$. We have $\gamma_a = 274n$, which is about 9268 when $n = 1373$. So a rational attacker would have to value breaking the specific user’s anonymity at at least \$92.68 to receive positive utility from attacking the protocol. In comparison, in January, 2006, this value would have been just \$18.56 showing how a recurring fee strategy can leverage an increase in the system’s popularity to deter attackers, while the current policy of a one-time fee remains constant.

An attacker who is satisfied by compromising any one user’s anonymity — perhaps to try to show that RFOR’s anonymity protection is limited — has a $B_{3,3}^{\text{any}}$ objective. We have $\gamma_a = \frac{2n+3}{(1-(1-(\frac{n}{3(n-1)})^3)^n)}$, which at $n = 1373$ is about 2749. A rational attacker with the goal of simply breaking anyone’s anonymity would need to value the goal at \$27.49 or more to profitably attack.

The attacker who values equally any information she receives about who is communicating with whom has the $B_{3,3}^{\text{all}}$ objective and has a far easier task. Here, $\gamma_a = \frac{27}{4}$, which does not depend on the number of participants at all. Such an attacker only needs to value the attack at about \$0.07 to profit from attacking, even if many more Tor routers join the network.

4 Entry Fees

We require only proof that each identity has paid a recurring fee, and we do not require proof that each identity is actually a separate entity. Moreover, the fee does not have to be paid to the administrator or other participant in the application. We need only ensure that some real cost has been provably paid before participating. Peers may pairwise prove to one another that they have paid the recurring fee each round; however, we expect in practice, a central trusted authority is likely to be used, just as Tor uses a trusted directory server to learn of other peers.

Micropayments [7, 8, 41, 42] can be used to purchase certificates valid for a certain number of minutes or rounds in one or more applications. The seller of such certificates has a much easier task than a certification authority: she does not have to verify the identity of the purchasers, prevent customers from purchasing multiple certificates, or prevent certificates from being transferred.

CAPTCHAs [43] are automated puzzles in widespread use that attempt to force human effort by using computer generate puzzles which are difficult for a computer to

solve, but easy for a human to solve. It takes the author an average of three seconds to solve and enter the type of CAPTCHAs used on sites such as `mail.com` and `yahoo.com`; this is equivalent to a cost of about \$0.01 at the average US individual wage (see <http://factfinder.census.gov>). Wages in other countries and economies of scale could drive these costs down significantly.

Another option for recurring fees is to use SMS messages. To apply this recurring fee, an SMS message is sent to the phone every application round, and no two identities can share the same phone number. A survey of current US cell phone plans reveals that most charge \$0.05 to receive a text message; though some plans that allowed unlimited reception would break this approach, reducing granularity to a monthly recurring fee. The interesting aspect of this approach is that the large monthly charges for a phone line are a deterrent only if it is purchased specifically to enable Sybil attacks. Obtaining multiple phone lines has little utility for users, so Sybils incur an extra charge. This illustrates that the networked application itself does not need to receive payment; we require only that the application can generate a cost that is incurred by the identity.

For RFOR, SMS is the easiest solution to implement, while micropayments are the most robust. We realize that anonymous communication systems are all volunteer networks and these real, recurring costs would diminish participation in the network, but our goal is to show how to better defend the system against the Sybil attack.

Many schemes for charging of one-time fees cannot be converted to recurring fees. For example, many past works have suggested the use of computation or storage as methods of imposing one-time fees (e.g., Abadi et al [2]). The real costs of these schemes is a diminished availability of the user's CPU, disk, or memory resource — a one-time purchase of additional hardware can replace these costs.

5 Related Work

Prevention of the Sybil attack has been discussed as part of the design of many distributed applications and protocols. Many follow Douceur's work and suggest prevention using a central authority, but several other approaches have been proposed, which we review below. We believe our work is the first to consider the economics of Sybil attackers in a general context using an economic analysis and we offer the most detailed analysis of Tor's and RFOR's vulnerability.

Before this broad review, we note other work related to our contributions and context. We assume that participants in p2p networks are rational agents. Shneidman and Parkes [39] give evidence of self-interested behavior in p2p applications. We also use ideas from game theory; Osborne and Rubinstein [34] give a rigorous introduction. In our previous work [26], we suggested a method of Sybil detection based on recurring fees. Finally, we are not the first to apply a cost-benefit analysis to security problems; e.g., See Meadows [29].

- **Trusted certification** [17, 22, 28, 32] Trusted certification is the most popular response to the Sybil attack. It is the only approach that has the potential to completely eliminate Sybil attacks. However, the certifying authority must ensure that each identity corresponds to exactly one entity, which may be costly for large-scale systems. To

prevent all Sybil attacks, the certifying authority must also ensure that no certificates are lost or stolen, which is probably impossible in almost all applications.

- **Reputation Systems** have often been suggested as a solution to the problem of Sybil attacks. Cheng and Friedman [12] classify reputation as symmetric or asymmetric. In symmetric reputation systems [18,33,33] an identity’s reputation depends solely on the topology of the trust graph, not on the relative positions in the trust graph of the identity and its querier. Cheng and Friedman prove formally that such reputation systems are susceptible to Sybil attacks. In asymmetric reputation systems [9, 18, 20], a trusted node determines on the reputation of all other nodes and Cheng and Friedman show the limited conditions under which Sybils are prevented. Unfortunately, asymmetric reputation systems inevitably penalize newcomers, who must prove themselves by offering benefits before getting anything in return.

- **Resource testing** [19,25,46] Resource tests include checks for computing ability, storage ability, and network bandwidth, as well as IP addresses. Both Freedman and Morris [19] and Cornelli et al. [14] suggest that requiring heterogeneous IP address (i.e., addresses in separate autonomous systems) is more effective at preventing Sybils than just requiring an IP address. Similarly, the SybilGuard technique [46] probabilistically weeds out Sybil identities based on the structure of social network graphs. Edges between nodes are assumed to imply “strong trust” in the real-world, a much stronger implication than is typical in social networks. SybilGuard can be used when there is a significant overlap between real-world social networks and participants in an online application and when users can be trusted to follow edge trust rules. This limits its applicability. For example, the social networks captured by MySpace, Friendster, or the PGP key-signing tree would not contain valid edges.

- **Recurring fees** [4,18,21] These works are the closest to ours, in that they consider recurring, rather than one-time costs. Awerbuch and Scheidler [4] suggest the use of Turing tests such as CAPTCHAs to impose recurring fees, but do not do an economic analysis. Dragovic et al. [18] require certification of identities, but this certification is not trusted; rather, it is seen as a way of imposing identity creation costs. Gatti et al. [21] is the work most similar to ours; it uses an economic, game-theoretical approach to examine when attacks on censorship resistant networks are cost-effective.

Other approaches that we do not review here due to lack of space include the following: trusted devices [32, 36], which like PKIs must avoid duplication; verifiable auditing [3, 40], for example by asking for the factors of a large number; physical observation [10,35], which are typically proposed for mobile computing and do not entail a recurring cost for the attacker.

6 Conclusions

In this paper, we evaluate Sybil attacks from an economic point of view. We define the Sybil valuation as a way of quantifying the relative strength of attackers and use it as a quantitative measure of the application robustness. Our results show that the susceptibility to Sybil attacks varies considerably, and can vary for different attacks, as we examined for the Tor network. We show that, in contrast to one-time fees, recurring per-identity entry fees can discourage Sybil attacks in many cases by ensuring a cost

for the attacker that is linear with the number of participants. These results provide a more fine-grained understanding of the attack and allow protocol designers to measure the effectiveness of defenses, which is important since the attack is difficult to prevent using standard computer security measures.

References

1. Department of state bureau of diplomatic security: Investigating passport and visa fraud. <http://www.state.gov/m/ds/investigat>.
2. M. Abadi, M. Burrows, M. Manasse, and T. Wobber. Moderately Hard, Memory-Bound Functions. *Trans. Inter. Tech.*, 5(2):299–327, 2005.
3. K. Anagnostakis and M. Greenwald. Exchange-Based Incentive Mechanisms for Peer-to-Peer File Sharing. In *Proc. ICDCS*, pages 524–533, Mar. 2004.
4. B. Awerbuch and C. Scheideler. Group Spreading: A Protocol for Provably Secure Distributed Name Service. In *Proc. ICALP*, pages 183–195, July 2004.
5. R. Bhattacharjee and A. Goel. Avoiding Ballot Stuffing in eBay-like Reputation Systems. In *Proc. Wkshp on Econ of P2P Systems*, pages 133–137, August 2005.
6. M. Bianchini, M. Gori, and F. Scarselli. Inside PageRank. *Trans. Inter. Tech.*, 5(1):92–128, 2005.
7. M. Blaze et al. TAPI: Transactions for Accessing Public Infrastructure. In *Proc. IFIP-TC6 Intl Conf Personal Wireless Communications*, pages 90–100, Sept. 2003.
8. M. Blaze, J. Ioannidis, and A. Keromytis. Offline Micropayments without Trusted Hardware. In *Proc. Fin. Crypto.*, pages 21–40, Feb. 2001.
9. S. Buchegger and J.-Y. L. Boudec. A Robust Reputation System for P2P and Mobile Ad hoc Networks. In *Proc. Wkshp on Econ of P2P Systems*, 2004.
10. S. Capkun, J. Hubaux, and L. Buttyan. Mobility helps peer-to-peer security. *IEEE Trans. Mobile Comp.*, 5(1), Jan 2006.
11. G. Casella and R. Berger. *Statistical Inference*. Wadsworth, 2000.
12. A. Cheng and E. Friedman. Sybilproof Reputation Mechanisms. In *Proc. Wkshp on Econ of P2P Systems*, pages 128–132, August 2005.
13. A. Clausen. Online Reputation Systems: The Cost of Attack of PageRank. Master’s thesis, Univ. of Melbourne, 2003.
14. F. Cornelli, E. Damiani, and S. Samarati. Implementing a Reputation-Aware Gnutella Servent. In *Proc. IPTPS*, pages 321–334, Mar. 2002.
15. L. Cox and B. Noble. Pastiche: Making Backup Cheap and Easy. In *Proc. OSDI*, pages 285–298, Dec. 2002.
16. R. Dingledine, N. Mathewson, and P. Syverson. Tor: The Second-Generation Onion Router. In *Proc. USENIX Security Symp*, pages 303–320, Aug. 2004.
17. J. Douceur. The Sybil Attack. In *Proc. IPTPS*, pages 251–260, Mar. 2002.
18. B. Dragovic, E. Kotsovinos, S. Hand, and P. R. Pietzuch. Xenotrust: Event-based Distributed Trust Management. In *Proc. Intl Wkshp on Database and Expert Systems Applications*, page 410, 2003.
19. M. J. Freedman and R. Morris. Tarzan: A Peer-to-Peer Anonymizing Network Layer. In *Proc. CCS*, pages 193–206, Nov. 2002.
20. Y. Fu, J. Chase, B. Chun, S. Schwab, and A. Vahdat. SHARP: An Architecture for Secure Resource Peering. In *Proc. SOSP*, pages 133–148, Oct. 2003.
21. R. Gatti, S. Lewis, A. Ozment, T. Rayna, , and A. Serjantov. Sufficiently Secure Peer-to-Peer Networks. In *Proc. Wkshp on Econ of P2P Systems*, May 2004.
22. K. Hildrum and J. Kubiatowicz. Asymptotically efficient approaches to fault-tolerance in peer-to-peer networks. In *Proc. Intl Symp on Distributed Computing*, pages 321–336, 2003.
23. Y.-C. Hu, A. Perrig, and D. Johnson. Ariadne: A Secure On-Demand Routing Protocol for Ad hoc Networks. *Wireless Networks*, 11(1–2):21–28, 2005.
24. S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proc. Intl Conf on World Wide Web*, pages 640–651. Press, 2003.

25. P. Maniatis, D. S. H. Rosenthal, M. Rousopoulos, M. Baker, T. Giuli, and Y. Muliadi. Preserving Peer Replicas by Rate-Limited Sampled Voting. In *Proc. SOSP*, pages 44–59, 2003.
26. N. B. Margolin and B. N. Levine. Informant: Detecting Sybils Using Incentives. In *Proc. Fin. Crypto. (FC)*, February 2007.
27. N. B. Margolin and B. N. Levine. Quantifying resistance to the sybil attack. Computer Science Technical Report 2007-64, University of Massachusetts Amherst, December 2007.
28. S. Marti and H. Garcia-Molina. Limited reputation sharing in p2p systems. In *Proc. 5th conference on Electronic commerce*, 2004.
29. C. Meadows. A cost-based framework for analysis of denial of service in networks. *J. Comput. Secur.*, 9(1-2):143–164, 2001.
30. S. J. Murdoch. Hot or Not: Revealing Hidden Services by their Clock Skew. In *ACM Conference on Computer and Communications Security (CCS)*, pages 27–36, October 2006. <http://www.cl.cam.ac.uk/~sjm217/talks/ccs06hotornot.pdf>.
31. S. J. Murdoch and G. Danezis. Low-Cost Traffic Analysis of Tor. In *Proc. IEEE Symp on Security and Privacy*, pages 183–195, May 2005.
32. J. Newsome, E. Shi, D. Song, and A. Perrig. The Sybil Attack in Sensor Networks: Analysis & Defenses. In *Proc. IPSN*, pages 259–268, 2004.
33. N. Ntarmos and P. Triantafillou. SeAI: Managing Accesses and Data in Peer-to-Peer Sharing Networks. In *Proc. IPTPS*, pages 116–123, 2004.
34. M. J. Osborne and A. Rubinstein. *A Course In Game Theory*. MIT Press, 1994.
35. C. Piro, C. Shields, and B. N. Levine. Detecting the Sybil Attack in Ad hoc Networks. In *Proc. SecureComm*, pages 1–11, Aug 2006.
36. R. Rodrigues, B. Liskov, and L. Shrira. The design of a robust peer-to-peer system. In *Proc. SIGOPS European Wkshp*, Sep 2002.
37. B. Schneier. *Applied Cryptography*. John Wiley & Sons, 1996.
38. Seti@home. <http://setiathome.ssl.berkeley.edu>.
39. J. Shneidman and D. C. Parkes. Rationality and Self-Interest in Peer-to-Peer Networks. In *Proc. IPTPS*, pages 139–148, 2003.
40. M. Srivatsa and L. Liu. Vulnerabilities and Security Threats in Structured Overlay Networks: A Quantitative Analysis. In *Proc. ACSAC*, pages 252–261, Dec 2004.
41. A. Stavrou, D. L. Cook, W. G. Morein, A. D. Keromytis, V. Misra, and D. Rubenstein. WebSOS: An overlay-based system for protecting web servers from denial of service attacks. *J. Comm Networks*, 48(5), August 2005.
42. A. Stavrou et al. A Pay-Per-Use DOS Protection Mechanism for the Web. In *Proc. Applied Cryptography and Network Security Conf (ACNS)*, June 2004.
43. L. von Ahn, M. Blum, N. Hopper, and J. Langford. CAPTCHA: Using Hard AI Problems for Security. In *Proc. Eurocrypt*, pages 294–311, 2003.
44. M. Wright, M. Adler, B. Levine, and C. Shields. The predecessor attack: An analysis of a threat to anonymous communications systems. *TISSEC*, 7(4):489–522, 2004.
45. M. Wright, M. Adler, B. N. Levine, and C. Shields. Passive-Logging Attacks Against Anonymous Communications Systems. *TISSEC*, 11(2), May 2008.
46. H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. SybilGuard: Defending Against Sybil Attacks via Social Networks. In *Proc. SIGCOMM*, pages 267–278, Sept. 2006.
47. M. Yurkewych, B. N. Levine, and A. L. Rosenberg. On the Cost-Ineffectiveness of Redundancy in Commercial P2P Computing. In *Proc. CCS*, pages 280–288, 2005.