Quantifying and Discouraging Sybil Attacks

N. Boris Margolin and Brian N. Levine

Abstract

The Sybil attack remains a largely open problem in peer-to-peer networking. Treating attackers as economically rational entities, we quantify the cost-benefit analysis of launching Sybil attacks. We propose the *equilibrium valuation* measure, a quantitative measure of the cost of achieving specific attacker objectives in the presence of entry fees for each identity. This measure allows protocol designers to set fees that discourage the maximum possible number of attackers. The cost of the Sybil attack is constant when an application uses one-time entry fees such as resource tests. However, our analysis shows that, for many attack objectives, the cost of the Sybil attack is linear in the number of users when recurring fees are used. Recurring entry fees can take many forms in practice, including CAPTCHAs, SMS messages, and anonymous electronic cash.

Our analysis is broad and can be applied to a number of existing p2p applications. As an example, we apply the equilibrium valuation measure to the TOR anonymous routing network, which currently has around 220 peers. We find that a Sybil attacker attempting to break a TOR user's anonymity must value that outcome at 1,485 times the recurring cost of acquiring an identity. If recurring fees are $0.01 per path reformation, only attackers with a valuation over $14.85 for breaking the user's anonymity can expect to profit from a Sybil attack. The valuation will increase linearly with the recurring fee or with the number of users, demonstrating quantitatively the risk the Sybil attack poses to TOR and similar p2p networks.

1 Introduction

In the Sybil attack, introduced by Douceur [8], a malicious entity creates many counterfeit identities and uses them to launch a coordinated assault on a peer-to-peer (p2p) application. The attack is applicable to many important p2p applications including anonymous routing [7], file storage [5], and mobile ad hoc networking [**?**], and cannot be prevented using resource tests in most situations. The attack has been widely studied [20, 16, 14, 11, 6] but remains unsolved in general, and to our knowledge no concrete measure of resistance of applications to the Sybil attack has been proposed. Addressing the general Sybil attack, we make the following contributions:

- We present the *valuation ratio*, a measure of an attacker's interest in achieving her specific objectives.
- We present the *equilibrium valuation*, a measure of the cost of achieving a specific objective using a Sybil attack, in terms of the cost of acquiring identities. Attackers only profit (on average) when their valuation ratio exceeds the equilibrium valuation.
- We show that protocols that charge a recurring fee per participating identity can be an effective disincentive against successful Sybil attacks. We also show that protocols based on a one-time fee per participating identity have limited effectiveness.

Our analysis can be applied to numerous p2p systems. As an example, we evaluate TOR [7], an open-source anonymous routing system that has 220 proxies. If TOR used a recurring entry fee, a rational entity interested in a particular user's communications would have to value knowledge of a single connection at 1,485 times the entry fee in order to launch a Sybil attack. At a fee setting of $0.01, this is $14.85, which may be enough to discourage many casual attackers; for TOR users more concerned about their anonymity, the fee could be set higher. Entry fees are

E	Set of entities
I	Set of identities
S_e	Strategies of an entity
σ_q	Strategy of Sybil attack of a certain size
O	Set of outcomes
u_e	An entity's utility function
$\tau_e(o)$	An entity's raw utility function
$\pi_e(o)$	An entity's cost utility function
A	A set of objectives
$B^{\text{spec.}}_{c,k}$	Binomial objective against specific user
$B^{\text{any}}_{c,k}$	Binomial objective against any users
$B^{\text{all}}_{c,k}$	Binomial objective against all users
$H_{c,k}$	Hypergeometric objective
$V_{\alpha,\beta}$	Voting objective
T_c	Infinite time horizon objective
O^r_A	Outcomes with specified number of successes
ι_a	Function counting objective success in an outcome
λ_m	Entity's valuation of an objective in terms of the fee
γ^*_A	Equilibrium valuation of an objective
q	Number of Sybil identities controlled by an entity
n	Number of identities not controlled by the entity

Table 1: Variables used

not necessarily monetary, but can take many forms, including resource tests, CAPTCHAs [23], IP addresses, and SMS messages.

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 application entry fees. This analysis defines a valuation ratio for users and attackers, and allows us to calculate the the equilibrium valuation for several different p2p protocols. We also discuss a number of different recurring and non-recurring entry fee types. In Section 4 we discuss related work, and we offer concluding remarks in Section 5.

Appendix A gives details of the equilibrium valuation results presented in Section 3.

2 Model

Our model of a p2p computing environment is based on Douceur's work [8] with extensions to include the utility functions and strategies of entities.

2.1 Protocol

In our model, each peer participating in a p2p protocol is a unique *identity* that is controlled by a rational actor [15] known as an *entity*. An entity launches a Sybil attack when it secretly controls multiple identities in order to achieve its objectives.

Identities send messages to each other through a *communications cloud*. The communication cloud precludes definite identification using direct observation. Like Douceur, we assume that entities have a reasonable level of computational power, so public-key cryptography can be used. In general, we assume that all messages are signed

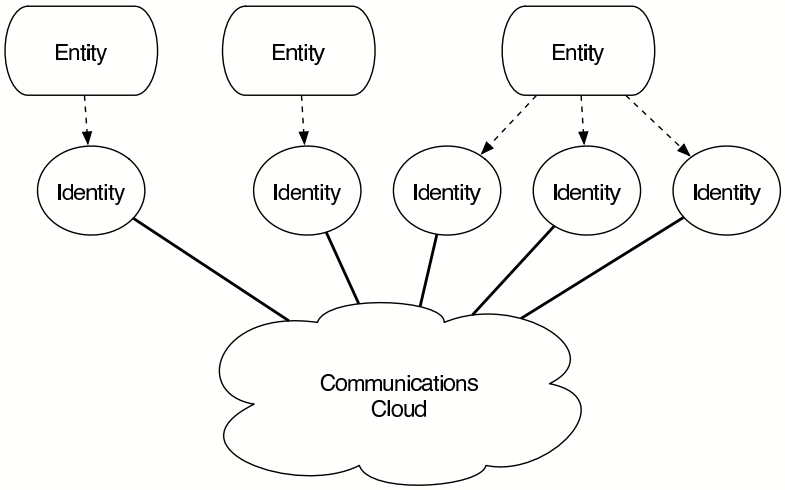

Figure 1: Our Model of p2p Applications

by identities using public-key cryptography; their unique identifier within the protocol is simply their public key.[1] This model is illustrated in Figure 1.

We model p2p protocols as having an *entry phase* and a *service phase*. (To simplify the analysis we assume the service phase is the same for all identities.) We assume that all identities participating in the service phase have executed the entry phase properly.[2] Protocols may impose an *entry fee* during the entry phase, which may be in the form of money, computational puzzles, CAPTCHAs [23], or many other forms. Peers may be forced to repeat the entry phase (and fee) after one or more service phases.

2.2 Entity Utility

In our model, entities are rational actors. This means that they have a specific utility for each possible protocol outcome, and that they apply strategies that give them the highest possible expected utility. For example, a hacker who controls many computers in different locations is a single entity, since she has a single utility function. Rational actors perform a cost-benefit analysis to determine what action to take — including whether or not to launch a Sybil attack against a specific protocol.

Our model follows basic game theory [15]. 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 consider. Entities must decide on a single strategy based on their knowledge and goals.[3] For example, a Sybil attack with a certain number of identities is a strategy an attacker might choose.

When the protocol is deterministic, the combination of the strategies of participating entities completely defines an *outcome*. Non-deterministic protocols include the effects of chance in the form of the actions of an additional entity R. S_R is the set of strategies for this "random player". (S_R is not a rational agent; it "chooses" its strategy according to some probability distribution.) An outcome $o \in O$ is a selection of one strategy from each of these sets, so o is a tuple

$$(s_{e_1}, s_{e_2}, \cdots, s_{e_n}, s_R)$$

drawn from

$$O = S_{e_1} \times S_{e_2} \times \cdots \times S_{e_n} \times S_R \tag{1}$$

[1] These public keys are *not* part of any PKI, and so are not securely associated with any identifying data about an individual.

[2] Proof that an identity has executed the entry phase properly comes in the form of direct interaction with each other identity, as for example with mutual puzzle solving or certification from one or more trusted protocol entry agents.

[3] We do not consider mixed strategies.

where $n = |E|$.

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 consists of a *raw utility* $\tau_e(o)$ and a *cost utility* $\pi_e(o)$ determined by payments made by e in outcome o:

$$u_e(o) \equiv \tau_e(o) + \pi_e(o) \tag{2}$$

where $\tau_e : O \to \mathbb{R}$ and $\pi_e : O \to \mathbb{R}$. In our analysis below, the cost, $\pi_e(o)$, will be the product of the entry fee and the number of identities controlled by the entity.

In many cases, we will be concerned with expected utility. An entity e_k's expected utility given a choice of strategy $x \in S_{e_k}$ is

$$\mathrm{E}[u_{e_k}|s_{e_k} = x] = \sum_{o \in O} u_e(o)\Pr[o|s_{e_k} = x], \tag{3}$$

where o is the tuple $(s_{e_1}, \ldots, s_{e_k}, \ldots, s_{e_n}, s_R)$, with s_{e_i} representing the strategy chosen by the entity e_i out of the set S_{e_i}.

2.3 Sybil Attacks

For an attacker 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. (The degenerate cases σ_0 and σ_1 correspond to not participating at all, or to entering a single identity.)

Let A be the set of the objectives (such as controlling an entire anonymity path) that an attacker can attempt to achieve using Sybil attacks. Each $a \in A$ partitions the set of outcomes, O, into those outcomes in which the objective is met and those in which it is not: O_a^0 and O_a^1, respectively. When it is possible for the objective to be reached repeatedly (i.e., against several victims at once), the objective partitions O into sets O_a^r, where r represents the number of successful attacks. We further define an *objective success count* operator ι_a, which gives the number of successes by m in the outcome.

$$\iota_a(o) \equiv r \text{ such that } o \in O_a^r. \tag{4}$$

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

$$\tau_m(o) = v\iota_a(o) \tag{5}$$

The attacker's expected utility from launching a Sybil attack with q identities is given

$$\mathrm{E}[\tau_m|s_m = \sigma_q] = \sum_{o \in O} v_m \iota_a(o)\Pr[o|\sigma_q]. \tag{6}$$

We assume that honest entities do not gain any raw utility from Sybil attacks; thus for each honest entity $h \in E$,

$$\mathrm{E}[\tau_h|s_h = \sigma_1] \geq \mathrm{E}[\tau_h|s_h = \sigma_i] \text{ for all } i \geq 2. \tag{7}$$

2.4 Objectives

Entities launch attacks in order to achieve some *objective*. We model the objectives as control of some key identities during the service phase of the protocol, and assume that if attackers achieve this control, they will correctly perform whatever additional actions necessary to reach their ultimate goal. (If control of these key identities gives a probability rather than certainty of achieving some ultimate goal the analysis is more complicated but not significantly changed.)

Many objectives are possible; we analyze four general types of objectives that are widely applicable: *binomial*, *hypergeometric*, *voting*, and *infinite-horizon* objectives. Example p2p applications subject to each type of objective are shown in Figure 2.

Objective	Type	Applications
$B_{c,k}$	Binomial	TOR [7], Predecessor [24]
$H_{c,k}$	Hypergeom.	SETI@HOME, Pastiche [5]
$V_{\alpha,\beta}$	Voting	Byzantine Agreement [12]
T_c	Infinite Horiz.	Any of the above, when Sybil attacks can be repeated indefinitely

Figure 2: Four General Objectives

In some p2p protocols, fixed-size subgroups of identities are created of size k. For example, anonymous routing protocols will create paths of k peers, while Pastiche [5] will store back up data from a source node with k other peers. In binomial objectives, k identities are chosen *with* replacement[4] from I. In hypergeometric objectives, k identities are chosen *without* replacement from I. In both objectives, the attacker desires to control c of the k identities chosen.

Two other objectives do not consider subgroups. In voting objectives, there is no subgroup and the attacker attempts to control some fraction of I. Infinite horizon objectives allow repeated chances for success without additional cost, as we explain below.

In the case of Hypergeometric, Voting, and Infinite-Horizon objectives, and two of three subtypes of Binomial objectives, the raw utility function, τ_m, for an attacker m partitions all protocol outcomes into failure, O_a^0, and success, O_a^1, in achieving the objective. In the All Identity subtype of Binomial objectives, τ_m partitions protocol outcomes into classes O_a^0 to O_a^N, where N is the number of identities not controlled by the attacker.

• **Binomial** objectives succeed when attackers control c of k specific identities in a group. Such objectives are applicable to the many applications in which identities use peer groups of size k, chosen with replacement, to accomplish their goals. For instance, in anonymous routing protocols, such as TOR [7], each identity use a sequence of other identities to route messages.

When each identity has its own group, a malicious entity can attack the group of a *specific* individual, the group of *any* individual, or the groups of *all* individuals.

– **Specific Identity Objectives.** In this case, the attacker's utility is proportional to the probability of success against the one identity. Such objectives are denoted $B_{c,k}^{\text{spec.}}$.

– **Any Identity Objectives.** The attacker may want to achieve control of a victim's peer group, without caring who the victim is. For instance, a hacker may want to prove that she can break a certain application. In this case, the attacker's utility is proportional to the probability of success against any identity. Such objectives are denoted $B_{c,k}^{\text{any}}$.

– **All Identity Objectives.** Finally, an attacker may want to cause as much damage as possible, or to learn about as many individuals as possible. In this case, the attacker's utility is proportional to the total number of group control successes against all identities in the protocol. Such objectives are denoted $B_{c,k}^{\text{all}}$.

• **Hypergeometric** objectives are similar to binomial objectives, but identities are chosen without replacement. Such objectives are denoted $H_{c,k}$. SETI@home [5] and Pastiche [5] are subject to the hypergeometric objectives, since identities in peer groups are chosen without replacement, for redundancy.

• **Voting** objectives succeed when an attacker can control some fraction α of the nodes in the system, plus some additional identities β. Voting objectives are denoted $V_{\alpha,\beta}$. For instance, the OM algorithm for the Byzantine generals problem [12] can be successfully attacked by $\frac{N}{3}+1$ Sybil identities (i.e., traitors).

[4] Commonly, selection of identities is uniformly at random, and we assume so here for all objectives.

[5] `http://setiathome.ssl.berkeley.edu/`

• **Infinite Horizon** objectives are a broad class of objectives in which the attacker can launch Sybil attacks repeatedly without additional cost — as when entry fees are charged only one time per identity. (Section 3.4 discusses different types of entry fees.) Any of the above protocols may have Infinite Horizon objectives under these conditions. We denote such objectives as T_c. Since an attack with *any* probability of success is certain to succeed eventually, a given strategy σ_q has either no chance of success, or is guaranteed success. The only strategies which the attacker needs to consider are σ_0, entering no identities, and σ_c, entering the minimum number of identities required for success, and we have

$$(s_{e_1}, \ldots, s_m, \ldots, s_{e_n}, s_R) \in O_a^1 \text{ if } s_m = \sigma_c \tag{8}$$

and

$$\mathrm{E}[\tau_m | s_m = \sigma_c] = v. \tag{9}$$

3 Discouraging Sybils

An entity attempts a Sybil attack to achieve a valued objective. Using these measures, we show that application designers can raise the cost of Sybil attacks by raising the entry fee, which discourages attackers who find the cost of the attack too high for the raw utility of the outcome. However, if the entry fee is too high, an unacceptable number of honest participants will drop out. So the relative strength of the attacker and defender of a protocol depends on the attacker's valuation of its objectives relative to honest users' valuation of the protocol.

In this section, we present a measure for the level of interest an entity has in its objective, as a multiple of the entry fee. We call this measure the *valuation ratio* and denote it λ_m, where $m \in E$ is the attacker.[6] The higher this measure, the more determined the attacker is to achieve her objective, and the more difficult it is for a defender to discourage her.

The difficulty of achieving a particular objective depends on the nature of the objective itself, not on the preferences of an attacker. For an objective a, we define the *equilibrium valuation*, denoted γ_a^*; this represents the number of identities an attacker must enter to achieve her objectives, normalized by the probability of success of the attack. When an attacker has a valuation ratio exceeding the equilibrium ratio, it is profitable on average for her to launch a Sybil attack; when she has a valuation ratio below the equilibrium ratio, she will lose money on average by launching such an attack.

Using these measures, we show that the resistance of protocols to Sybil attacks varies considerably according to the objective of the attacker. Infinite Horizon objectives are especially easy to achieve. When they can be prevented by the defender, by charging repeating rather than one-time fees, all but the most determined attackers can be discouraged from attacking many classes of p2p protocols.

3.1 Valuation Ratio and Equilibrium Valuation

In this section, we define λ_m, a measure of the interest entity m has in an objective, and γ_a^*, a measure of the difficulty of success in an objective $a \in A$. The first of these measures depends on the preferences of the attacker and the fees charged by the protocol, while the second depends only on the nature of the objective.

The *valuation* an entity has for a protocol is the maximum fee it is willing (or able) to pay each entry phase. To participate, an entity's valuation must be greater than equal to the product of the number of identities it enters and the entry fee.

By setting the entry fee f properly, an p2p protocol designer can discourage the greatest number of Sybil attacks possible without discouraging an unacceptable number of honest entities. Let $h \in E$ be an honest entity willing to pay for protocol participation. (Although the maximum fee will differ for different entities, we assume that h

[6] The objective is generally obvious from context.

represents a typical honest user of the protocol.) Rational entities will not participate if their utility is negative, and it is easy to show that the f must be less than or equal to $E(\tau_h)$.

$$\begin{aligned} \mathrm{E}[u_h] &\geq 0 \\ \mathrm{E}[\tau_h] + \mathrm{E}[\pi] &\geq 0 \\ \mathrm{E}[\tau_h] - f \times 1 &\geq 0 \\ \mathrm{E}[\tau_h] &\geq f \end{aligned} \tag{10}$$

Next we consider a malicious entity $m \in E$. Denote the maximum raw utility of an outcome is $\max v$, where

$$\max v \equiv \max_{o \in O} u_m(o). \tag{11}$$

Then, as an upper bound, the most identities, q, the attacker m will be willing to enter is $\max v/f$. We assume that the expected net utility of m is non-negative because m is rational.

$$\begin{aligned} E[u_m | s_m = \sigma_q] &\geq 0 \\ E[\tau_m | s_m = \sigma_q] + E[\pi_m | \sigma_q] &\geq 0 \\ E[\tau_m | s_m = \sigma_q] - fq &\geq 0 \\ \frac{E[\tau_m | s_m = \sigma_q]}{f} &\geq q. \end{aligned} \tag{12}$$

Since $E[\tau_m | s_m = \sigma_q] \leq \max v$, we find that the attacker does not profit from launching a Sybil attack with q identities unless

$$\frac{\max v}{f} \geq q. \tag{13}$$

If we know the objective $a \in A$ that m is attempting to meet, we can improve this requirement, and define precisely when a rational attacker prefers to make an attack, when she prefers not to, and when she is indifferent towards these options. The *equilibrium valuation* of a Sybil attack of size q is denoted γ_a^q and quantifies the requirements for a Sybil attack with q entities to be profitable; γ_a^* quantifies the requirements for *any* such attack aimed at achieving the objective to be profitable.

Note that

$$E[\tau_m | s_m = \sigma_q] = \sum_{o \in O} v \iota_a(o) \mathrm{Pr}[o | \sigma_q];$$

substituting this in Inequality (12) we have

$$q \leq \frac{v}{f} \sum_{o \in O} \iota_a(o) \mathrm{Pr}[o | \sigma_q] \tag{14}$$

and we see that the number of attackers entered, q, depends on the ratio v/f, normalized by the expected number of objective successes for various values of q. We call the ratio of v and f the attacker's *valuation ratio* and denote it as

$$\lambda_m \equiv \frac{v}{f}. \tag{15}$$

Solving Inequality (14) for λ_m, we find that a σ_q attack is profitable if and only if

$$\lambda_m \geq \frac{q}{\sum_{o \in O} \iota_a(o) \mathrm{Pr}[o | \sigma_q]} \tag{16}$$

Therefore, if an attacker, m, more highly values her objective (as indicated by her valuation ratio λ_m), then for her to launch a Sybil attack she requires a lower likelihood of success. Since m expects to profit from a Sybil attack of size

q when λ_m exceeds this value and expects to lose money when λ_m is less than this value, we call it the *equilibrium valuation*, denoted γ_a^q and defined

$$\gamma_a^q \equiv \frac{q}{\sum_{o \in O} \iota_m(o) \Pr[o|\sigma_q]}.$$

The attacker can control the number of identities q in order to minimize γ_a^q. This minimum value is denoted γ_a^*:

$$\gamma_a^* \equiv \min \gamma_a^q, \text{ where } q \geq 1 \tag{17}$$

This is the lowest possible value of γ_a^q for any q, so it represents the minimum λ_m an attacker must have in order to launch a Sybil attack of any size. Note that the quantity on the right side of Inequality 17 depends only on the characteristics of the protocol and of the attack, not on the valuation of the Sybil. A rational attacker m can benefit from a Sybil attack a against a protocol if and only if

$$\lambda_m \geq \gamma_a^* \tag{18}$$

Inequality 18 says nothing about the *resources* available to a particular attacker: she may value an objective highly, but lack the resources to enter enough Sybil identities 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.

3.2 γ_a^* for Four Types of Objectives

We now quantify γ_a^*, the susceptibility of protocols to Sybil attacks. The susceptibility of a protocol depends on the attackers objective (as discussed in Section 2.4), and each protocol only admits certain attacker objectives. We group our results by the type of objective and not by what type of service the protocol provides. In Section 3.3, we use TOR [7] as a concrete example and show in monetary terms how much an attacker must value an objective in order to launch an attack if the entry fee is a recurring \$0.01.

Table 2 has five columns of results for each objective discussed in Section 2.4: the binomial, hypergeometric, voting, and infinite horizon objectives. In the first column, **Min. q** denotes the absolute minimum number of identities that an attacker needs to achieve the objective. The second column, $\boldsymbol{\gamma_a^q}$, denotes the equilibrium valuation for a specific number of Sybil identities, q. The third column shows the **attacker's best q**, which is the number of identities that maximizes total utility for the attacker. The fourth column, $\boldsymbol{\gamma_a^*}$, shows the minimum difficulty of achieving the objective; while the final column, $\boldsymbol{\Theta(\gamma_a^*)}$ shows the asymptotic bound of γ_a^* as the number of participants in the protocol increases.

The derivations of each entry in Table 2 appear in the Appendix. We discuss the details of the Table 2 below.

• **The infinite horizon** objective, T_c, is easily achieved for low values of c, which is the number of identities required for positive probability of success. Specifically, $\lambda_{T_c}^{\max}$ is c. For example, if the underlying objective is a predecessor attack on an anonymity system requiring 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 — leading to a very inexpensive Sybil attack. In other words, this case shows that one-time fees are not well-suited to discouraging Sybil attacks.

• **Binomial** objectives vary in difficulty depending on if the objective has as its intended victim 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 between.[7]

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.

[7] Note there is no closed-form expression for γ_a^* for Binomial objectives with general c, k.

A	**Min. q**	γ_a^q	**Attacker's best q**	γ_a^*	$\Theta(\gamma_a^*)$
T_c	c	q	c	c	c
$B_{1,k}^{\text{spec.}}$	1	$q/(1-(\frac{n}{q+n})^k)$	1	$1/(1-(\frac{n}{n+1})^k)$	$k^{-1}n$
$B_{k,k}^{\text{spec.}}$	k	$q/(\frac{q}{q+n})^k$	$(k-1)n$	$\left(\frac{k}{k-1}\right)^{k-1} kn$	ekn
$B_{1,k}^{\text{any}}$	1	$q/(1-(\frac{n}{q+n})^{kn})$	1	$1/(1-(\frac{n}{n+1})^{kn})$	e^{-k}
$B_{k,k}^{\text{any}}$	k	$q/(1-(1-(\frac{q}{q+n})^k)^n)$	$\boldsymbol{(k-1)n+k}$	$\frac{(k-1)n+k}{(1-(1-(1-\frac{n}{(n+1)k})^k)^n)}$	$\boldsymbol{kn}$
$B_{1,k}^{\text{all}}$	1	$q/(n(1-(\frac{n}{q+n})^k))$	1	$1/n(1-(\frac{n}{n+1})^k)$	k^{-1}
$B_{k,k}^{\text{all}}$	k	$q/(n(\frac{q}{q+n})^k)$	$(k-1)n$	$\left(\frac{k}{k-1}\right)^{k-1} k$	k
$H_{1,k}$	*Same as $B_{1,k}^{spec.}$*				
$H_{k,k}$	k	$q\binom{n+q}{k}/\binom{q}{k}$	$\boldsymbol{(k-1)n+k}$	$\frac{(nk-n)!(nk+k)!}{(nk)!((n+1)(k-1))!}$	$\boldsymbol{kn}$
$V_{\alpha,\beta}$	$\lceil\frac{\alpha n+\beta}{1-\alpha}\rceil$	q	$\lceil\frac{\alpha n+\beta}{1-\alpha}\rceil$	$\lceil\frac{\alpha n+\beta}{1-\alpha}\rceil$	$n/(1-\alpha)$

T_c denotes an objective needing c Sybil identities for any positive probability of success and where there is only a one-time entry fee.

$B_{c,k}$ denotes the *Binomial* objective; the adversary's goal is to control c of k helper identities (chosen uniformly at random with replacement) for some specific identity, for any one identity, or for any identity.

$V_{\alpha,\beta}$ denotes the *Voting* objective. The attacker's goal is to control $\alpha N + \beta$ of the identities in the protocol.

$H_{c,k}$ denotes the *Hypergeometric* objective. The attacker's goal is to control c of k key identities in the protocol, chosen uniformly at random without replacement (usually based on some identifier associated with the identity.)

The values in bold face are conjectural, derived from numerical approximation.

Table 2: Summary of results for our analysis of four classes of objectives

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.

Figures 3, 4, and 5 offer a graphical comparison of these results. The figures show the effect of different attacker goals on the difficulty of achieving binomial objectives. When considering launching a $B_{1,3}^{\text{all}}$ Sybil attack with a binomial objective in which the attacker can benefit from each additional user attacked, the attacker does not even need to value compromising any single user at the level that honest users value the protocol for the attack to be profitable. On the other hand, as Figure 3 illustrates, to profit with a $B_{3,3}^{\text{spec}}$ objective considering a specific user, she must value the objective at $(k(k/k-1)^{k-1}) = 6.75$ times the number of honest users. For example, for a mixnets based anonymous routing protocol with $h = 300$ honest identities in the system, $\gamma_a^* = 2025$. Therefore, if an honest user pays \$0.01 to participate each round, the attacker must value a successful attack at \$20.30 to find positive utility in attacking the system.

• **Hypergeometric objectives** are similar to binomial 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.

• **Voting objectives**, $V_{\alpha,\beta}$, have equilibrium values increasing without bound (asymptotically linearly in $1/(1-\alpha)$) as α approaches 1. This is because the attacker, no matter how many identities she adds, can never control all identities in the protocol. Voting protocols with very high α values may not be very useful. For more customary

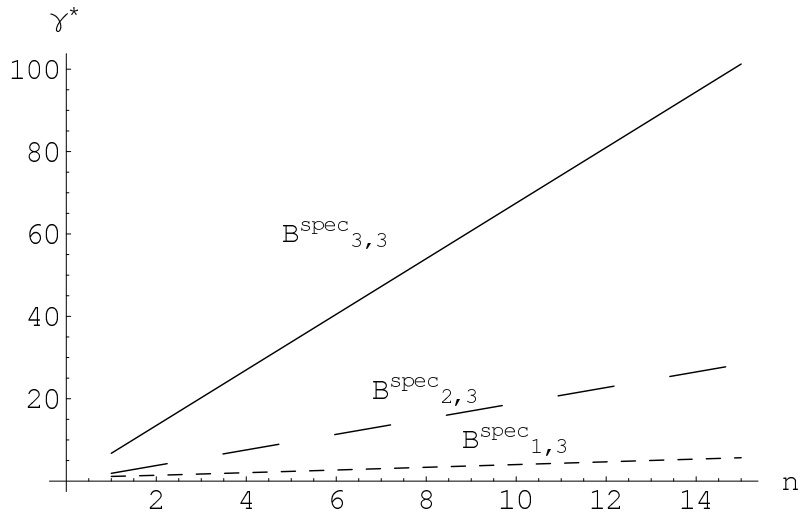

Figure 3: γ_a^* for $B_{c,3}^{\text{spec}}$, which is the binomial objective considering a specific user. Shown is $c = 1, 2$, and 3.

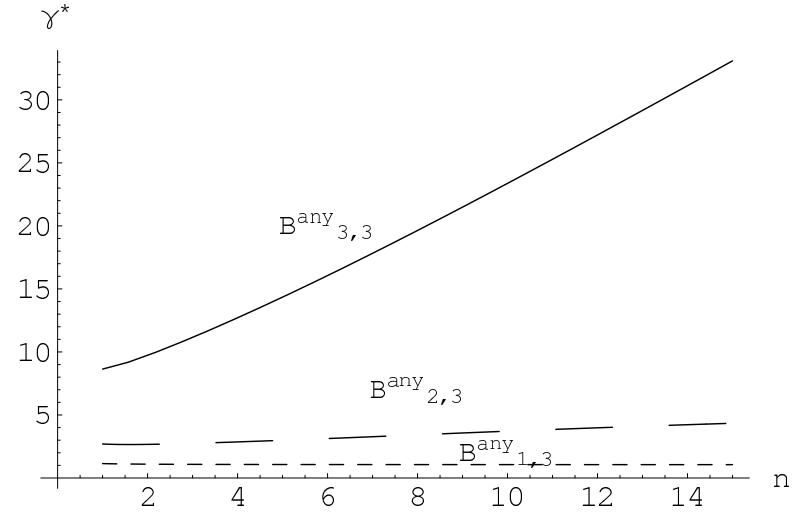

Figure 4: γ_a^* for $B_{c,3}^{\text{any}}$, which is the binomial objective considering any user. Shown is $c = 1, 2$, and 3.

50% + 1 voting ($\alpha = 0.5, \beta = 1$), γ_a^* is $h + 2$. For 2/3 super-majority voting it is $2h$.

3.3 Application: TOR Routing

In this section, we apply the equilibrium valuation measure to TOR [7]. TOR is a system of onion routers used for anonymity; circuit length is by default three (unless the user configures the system otherwise). In this section we assume that the path length is constantly 3, and that there are 220 TOR routers (as there are as of October 2005 [21]). We further assume that TOR users are charged a fee of \$0.01 every entry phase.

We analyze two objectives attackers of TOR might value:

- **Full-path objective:** attackers attempt to control all k nodes in the circuit, and can then know for certain that the endpoints are communicating.
- **Endpoints objective:** attackers attempt to control the first and last nodes of the circuit, and use a timing attacks [13] to determine if they are on the same circuit.

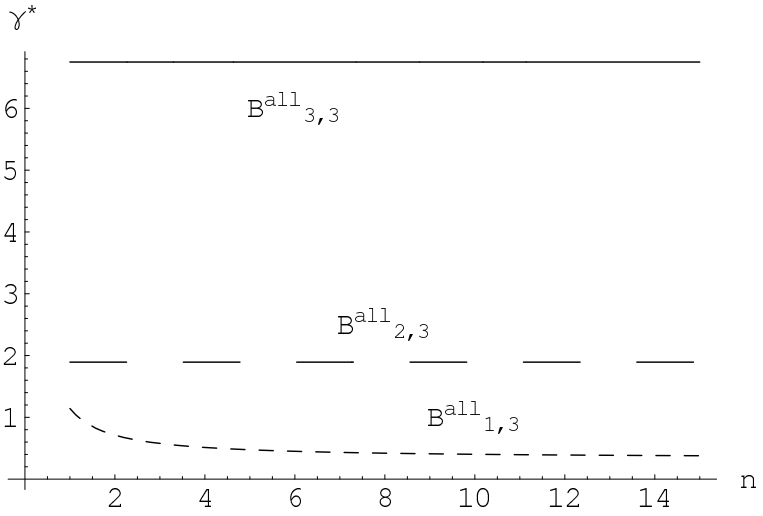

Figure 5: γ_a^* for $B_{c,3}^{\text{all}}$, which is the binomial objective considering all users. Shown is $c = 1, 2$, and 3.

3.3.1 Full-path objective

We first analyze the full-path objective considering a specific user. If there is a one-time fee of \$0.01 per identity, then the objective is an infinite-horizon objective T_3, and we have $\gamma_a^* = 3$. So an attacker would have to value the knowledge gained at \$0.03 or more for the attack to be profitable; a rational attacker who is not particularly determined is able to profit from an attack.

On the other hand, if the fee of \$0.01 is charged per circuit reformation, then the objective corresponds to $B_{3,3}^{\text{spec.}}$. We have $\gamma_a^* = \frac{27}{4}h$, which is 1,485 when $h = 220$. So a rational attacker would have to value breaking the specific user's anonymity at at least \$14.85 to receive positive utility from attacking the protocol.

An attacker who is satisfied by compromising any one user's anonymity — perhaps to try to show that TOR's anonymity protection is limited — has a $B_{3,3}^{\text{any}}$ objective. We have $\gamma_a^* = \frac{2h+3}{(1-(1-(1-\frac{h}{3(h-1)})^3)^h}$, which at $h = 220$ is about 443. A rational attacker with the goal of simply breaking anyone's anonymity would need to value this at \$4.43 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.

Unfortunately for users of TOR protocol, as the number of users n increases, the minimum required attacker valuation for the binomial objective considering a specific user or against any user grows only linearly. Similarly, the fee charged to users each entry phase also has only a linear affect on the growth of the attacker equilibrium valuation. However, the linear coefficient is relatively high at $27/4$. When the attacker's objective is to attack any user, her objective equilibrium valuation is constant regardless of the number of participants, and when her objective is to attack all users, her equilibrium valuation approaches zero as the number of TOR routers increases.

3.3.2 Endpoints objective

The objective of appearing as the endpoints in the circuit is similar to the Full-path objective but easier to achieve. With a one-time fee of \$0.01, the equilibrium valuation is 2, so an attacker only needs to value the objective at \$0.02 for the attack to be profitable when the objective is to compromise some specific user or any user; if the attacker values compromise of all users, her equilibrium valuation is less than one one-hundredth of a cent.

When the fee is charged per path reformation, the results are as follows. For the binomial objective, we have

Fee Type	One-time Cost	Recurring Cost
Electronic Cash	-	variable
CAPTCHA	-	$5.15 / person/hour
Computation	$100/GhZ	$0.01 / hr
Storage	$100 + $1.30/Gb	$0.01 / hr
Network	$100 + $20 / Mbps mo.	$0.04 / hr
VeriSign Certificate	$995/yr	$0.11 / hr
US Social Security #	$0 (legitimate) $2400/yr (bought) $30/14 mo (stolen)	$0 (legitimate) $0.27/hr (bought) $0.004/hr (stolen)

Figure 6: Entry Fee Types

against a specific user $\gamma_a^* = 4h$, or 880 at $h = 220$, which gives an attacker equilibrium valuation of $8.80. Against any user, we have $\gamma_a^* = \frac{h+2}{1-(1-(1-\frac{h}{2(h-1)})^2)^h}$, or about 222 at $h = 220$; so the equilibrium valuation is $2.20. Against all users, we have $\gamma_a^* = 4$. In this case the attacker only needs to value the objective at $0.04.

3.4 Entry Fees

In Section 3.2 we show that recurring fees are much more effective at discouraging Sybil attacks than one-time fees. With one-time fees, an attacker can enter just a few identities and repeat the attack each round until she succeeds in her attack, so they will only be discouraged when the fixed costs needed to obtain a non-zero chance of attack success exceed the return from a successful attack. With recurring fees, the attacker's expected return over a long period of time is a multiple of her return in each round; so if she does not have a positive return for the first round, she cannot achieve a positive return by attacking for multiple rounds. It is therefore often possible to discourage Sybil attacks when repeated costs are used.

In this section, we detail some of the ways that a protocol designer can impose per-identity costs on a protocol participant. We evaluate, informally, several types of entry fees and their rough cash equivalents to determine if they are practical as protocol devices. For each, we show how a protocol designer can use the fee type as a one-time charge of $1 per identity, and how a protocol designer can use the fee type as a recurring charge of $0.01 per identity per round.

As Figure 6 shows, entry fees have both a one-time and a recurring component, but usually either the one-time or the recurring component dominates.

Computation is the most convenient type of fee for a protocol designer to impose. There are a number of protocols, such as Dwork and Naor's [9] that can enable a group of identities to mutually solve computational puzzles in a way that assures each participant that the the other identities have paid the computational cost.

On the Dell website [8], a low-grade consumer computer (the Dell Dimension 4700) is currently available for around $100 per GhZ. Application designers can impose one-time costs of $\$C$ per identity by requiring users to solve puzzles that only a $\$\frac{C}{100}$ GhZ or faster processor can solve in the allotted time.[9]

A home computer system consumes roughly 100 watts [17]. At current prices, this is an impractical recurring fee.

- **One-time $1** Require solution within 10 seconds of a problem that would take a 1 GhZ processor 0.1 seconds.
- **Recurring $0.01** Require computations that are sufficient to fully load an average CPU for 1 hour.

[8] `http://www.dell.com/`

[9] We assume for convenience that clock rate is a reasonable measure of computational power.

Memory is another bound for computations. It is a more suitable fee than strict computational power, since the cost of memory is more linear than the cost of computational power. Abadi et al. describe some appropriate functions [1]. Dell currently sells memory for about $170 per GB.

- **One-time $1** Require solution within 10 seconds of a problem that would take a computer with 1 GB of memory 0.06 seconds.
- **Recurring $0.01** Require computations that are sufficient to completely fill an average computer's memory for 1 hour.

Storage fees require a user to verifiably store some amount of data per identity. Dell charges about $1.30 / GB for hard drive space. Users must have also access to a computer, adding about $100 to the total costs. Storage challenge fees are also primarily one-time fees.

- **One-time $1** Require users to store (and periodically verify) about 1 GB of data.
- **Recurring $0.01** Require storage sufficient to fill an average hard drive for 1 hour.

Bandwidth fees require a user to prove availability of a certain amount of bandwidth per identity, for instance by responding to a large number of challenges in a limited amount of time. The bandwidth trading website, *http://www.bandwidthmarket.com*, listed offers for $20 per Mbps per month in August 2005, which is about $0.03 per Mbps per hour. Bandwidth challenges are more effective recurring fees than storage or computational challenges. If the challenges for a single identity are intended to fill a fast, 10 Mbps consumer cable modem, the recurring costs are about $0.30 per challenge hour per identity.

- **One-time $1** Require users to show they control 50 kbps of bandwidth.
- **Recurring $0.01** If a round lasts r minutes, require users to continually transmit $2/r$ Mbps of data in addition to any application messages.

IP addresses can be obtained for about $30 per month.

- **One-time $1** Only allow 30 identities per round per IP address.
- **Recurring $0.01** Only allow 3000 single round uses of identities per month per IP address.

Electronic Cash is a nearly perfect type of recurring cost. Unfortunately, while there has been plenty of research on anonymous electronic cash, no system has been successfully deployed, so electronic cash is not usually a viable fee type. It is possible, though inconvenient, to make anonymous payments online; one way to do so is to set up an anonymous bank account at a site like Cardster [10] and link it to an online payment service such as *http://www.paypal.com*.

- **One-time $1** Require users to purchase reusable protocol entry certificates for $1 each, with only one usable at any given time.[11]
- **Recurring $0.01** Require users to purchase protocol entry certificates for $0.01 each that expire each round.

CAPTCHAs as described by Ahn et al. [23] are automated puzzles in widespread use that attempt to force some human effort by letting a computer generate puzzles which are difficult for a computer to solve, but easy for a human to solve. Such puzzles are usually based on humans' ability to read even distorted and obscured text. 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 [4].

[10] `http://www.cardster.net/`

[11] Users can use zero-knowledge techniques like Camenisch and Lysyanskays's [3] to prove that they have valid certificates without revealing them.

- **One-time $1** Give out reusable protocol entry certificates as a reward for solving 100 CAPTCHAs.
- **Recurring $0.01** Require users to solve one CAPTCHA per round.

Phone SMS Google (*http://www.google.com*) requires SMS messages to a unique cell phone number in order to register for an email account. (Users must also complete a CAPTCHA to register the account.) A survey of current US cell phone plans reveals that most charge $0.05 to receive a text message.[12]

- **One-time $1** Give users reusable protocol entry certificates in exchange for receiving 20 SMS messages.
- **Recurring $0.01** Give users a protocol certificate that expires in 5 rounds for every SMS message they receive.

Bank Accounts, used for example by Paypal [13], require users to demonstrate access to a bank account by reporting the timing and amount of small deposit made in to the account. Most bank accounts are not anonymous. It is possible to acquire more anonymous bank accounts (such as Cardster), but such accounts are relatively expensive to acquire. Bank account challenges primarily impose a one-time cost; it is usually possible to choose a type of bank account with no recurring fees.

- **One-time $1** Prove possession of a unique bank account in order to obtain some number of identities in a protocol.
- **Recurring $0.01** Not clearly applicable.

Identity Certificates are not generally considered fees, but rather proofs of legal identity. However, it is apparently possible to acquire a VeriSign identity certificate by paying a fee and presenting company letterhead. On the VeriSign website [14], the VeriSign recommended identity certificate, Secure Site Pro, is available for $995 per year, or $0.11 per hour. For those who want to use a p2p application for less than a year in total, requiring a VeriSign Secure Site Pro certificate per identity would impose significant one-time costs of $995 per identity. For those using the site for more than a year, the recurring cost would be $995 per identity per year.

- **One-time $1** Varies with certificate cost.
- **Recurring $0.01** Varies with certificate cost.

In the opinion of the authors, computation, memory, storage, bandwidth, CAPTCHAs, and phone SMS are the most practical current methods of imposing costs on users that want to join a p2p application. Both CAPTCHAs and SMS fees have a significant recurring component; for applications with access to a limited infrastructure, CAPTCHAs are the easiest to implement.

4 Related Work

Douceur's *The Sybil Attack* [8] first introduced Sybil attacks. Douceur showed that resource tests and vouching for identities have limited effectiveness in preventing Sybil attacks against p2p networks.

While the Sybil attack may seem theoretical, there is evidence from fields outside of computer science that simple Sybil attacks do occur. There are many instances of the use of multiple false identities in Internet polls [10], in chain letters [2], and in US elections [22].

The work that has followed Douceur can be divided into techniques for revealing common control of entities in special cases, and techniques for reducing the impact of Sybil attacks on protocols.

[12] Google also limits the number of mail accounts that can be created using one cell phone, so monthly and phone purchase fees would apply if an attacker attempted to create many identities over a long period of time.

[13] `http://www.paypal.com/`

[14] `http://www.verisign.com/`

Sybils can only be prevented completely using certification or some form of direct observation; the work on detecting Sybils has focused on direct observation in special types of networks. Newsome et al. [14] discuss Sybil attacks in sensor networks and propose several techniques for detecting Sybils. Radio resource testing and position verification use link-level information to check whether messages do come from two separate devices. (Of course, this does not preclude common control of the two devices.) Perrig et al. [16] discuss attacks on wireless networks; it is possible to defend against Sybil attacks quite successfully when the network topology is static. Our work does not address Sybil detection.

Certain types of attacks are more difficult to launch than others, and protocol designers can modify their algorithms to make Sybil attacks more costly for the attacker. Srivatsa and Liu [20] present several techniques for making successful Sybil attacks on DHT routing more difficult. Danezisis et al. [6] also address the Sybil attack in DHT networks. Assuming an honest bootstrap graph, "diversity routing" ensures that the impact of attacking nodes connected to any single bootstrap point is limited. In contrast, our work shows the effect of attack difficulty on attackers' cost-benefit analyses.

We believe that our work is the first to consider the entities behind Sybil attacks as rational agents. Most of the research identifies these attackers with physical devices, so the Sybil problem becomes linking identities with physical devices. Douceur did not identify entities with devices, but he also did not consider the rational agents; his work considers collaborating identities and Sybil identities interchangeable, although groups of collaborators differ from individual rational entities in their preferences.

We assume that participants in p2p networks are rational agents. Shneidman and Parkes [19] give evidence of self-interested behavior in p2p applications. We also use ideas from game theory; Osborne and Rubinstein [15] give a rigorous introduction.

5 Conclusions

In this paper, we evaluated the cost of Sybil attacks on p2p applications from an economic point of view. We defined the valuation ratio as a way of quantifying the relative strength of protocol attackers and the equilibrium valuation as a measure of the resistance of a protocol to attack. Our application of this measure suggests that the susceptibility of protocols to Sybil attacks varies considerably and recurring per-identity entry fees can discourage Sybil attacks in many cases.

These results provide a framework for understanding the Sybil attack. They have allowed us to quantify defenses against it, which is especially important since the attack is difficult to prevent using standard computer security measures.

References

[1] ABADI, M., BURROWS, M., MANASSE, M., AND WOBBER, T. Moderately hard, memory-bound functions. *ACM Trans. Inter. Tech. 5*, 2 (2005), 299–327.

[2] BULGATZ, J. *More Extraordinary Popular Delusions and the Madness of Crowds*. Three Rivers Press, 1992.

[3] CAMENISCH, J., AND LYSYANSKAYA, A. A signature scheme with efficient protocols. In *Proc. Intl Conf on Security in Communication Networks (SCN)* (2002), vol. 2576 of *LNCS*, Springer Verlag, pp. 268–289.

[4] U.S. Census. http://factfinder.census.gov.

[5] COX, L., AND NOBLE, B. Pastiche: Making backup cheap and easy. In *Proc. USENIX Symposium on Operating Systems Design and Implementation* (Dec. 2002).

[6] DANEZIS, G., LESNIEWSKI-LAAS, C., KAASHOEK, M. F., AND ANDERSON, R. Sybil-resistant DHT routing. In *Proc. ESORICS* (2005), pp. 305–318.

[7] DINGLEDINE, R., MATHEWSON, N., AND SYVERSON, P. Tor: The second-generation onion router. In *Proc. USENIX Security Symposium* (Aug. 2004).

[8] DOUCEUR, J. The Sybil Attack. In *Proc. Intl Wkshp on Peer-to-Peer Systems (IPTPS)* (Mar. 2002).

[9] DWORK, C., AND NAOR, M. Pricing via processing or combatting junk mail. In *Proc. Intl Cryptology Conference on Advances in Cryptology (CRYPTO)* (1993), pp. 139–147.

[10] JUDGE, P. ZDNet UK news: .net vote rigging illustrates importance of web services. http://news.zdnet.co.uk/software/0,39020381,2102244,00.htm, 2002.

[11] KARLOF, C., AND WAGNER, D. Secure routing in wireless sensor networks: Attacks and countermeasures. *Ad hoc Networks Journal (Elsevier) 1*, 2–3 (September 2003), 293–315.

[12] LAMPORT, L., AND FISCHER, M. J. Byzantine generals and transactions commit protocols. Tech. Rep. Opus 62, SRI Intl, Menlo Park, California, 1982.

[13] MURDOCH, S. J., AND DANEZIS, G. Low-cost traffic analysis of Tor. In *Proceedings of the 2005 IEEE Symposium on Security and Privacy* (May 2005), IEEE CS.

[14] NEWSOME, J., SHI, E., SONG, D., AND PERRIG, A. The Sybil attack in sensor networks: analysis & defenses. In *Proc. Intl Symposium on Information Processing in Sensor Networks (IPSN)* (2004), pp. 259–268.

[15] OSBORNE, M. J., AND RUBINSTEIN, A. *A Course In Game Theory*. MIT Press, 1994.

[16] PERRIG, A., STANKOVIC, J., AND WAGNER, D. Security in wireless sensor networks. *Commun. ACM 47*, 6 (2004), 53–57.

[17] ROUND, S. Microsoft press release: Computer power consumption tests. http://www.microsoft.com/nz/presscentre/articles/2001/august-01-consumption.aspx.

[18] SANZGIRI, K., DAHILL, B., LEVINE, B., AND BELDING-ROYER, E. A secure routing protocol for ad hoc networks, 2002.

[19] SHNEIDMAN, J., AND PARKES, D. C. Rationality and self-interest in peer to peer networks. In *Proc. Intl Wkshp on Peer-to-Peer Systems (IPTPS)* (2003).

[20] SRIVATSA, M., AND LIU, L. Vulnerabilities and security threats in structured overlay networks: A quantitative analysis. In *Proc. ACSAC* (2004), pp. 252–261.

[21] Number of running tor nodes. http://www.noreply.org/tor-running-routers/.

[22] VIGLUCCI, A., TANFANI, J., AND GETTER, L. Herald special report: Dubious tactics tilted mayoral votes. Miami Herald, February 8, 1998.

[23] VON AHN, L., BLUM, M., HOPPER, N., AND LANGFORD, J. CAPTCHA: Using hard AI problems for security. In *Proc. of Eurocrypt* (2003), pp. 294–311.

[24] WRIGHT, M. K., ADLER, M., LEVINE, B. N., AND SHIELDS, C. The predecessor attack: An analysis of a threat to anonymous communications systems. *ACM Trans. Inf. Syst. Secur. 7*, 4 (2004), 489–522.

A γ_a^* Calculations

In this appendix, we give the derivations of γ_a^* for the Infinite Horizon, Binomial, Hypergeometric, and Voting objectives. Recall that

$$\gamma_a^q \equiv \frac{q}{\sum_{o \in O} \iota_m(o) \Pr[o|\sigma_q]}.$$

and

$$\gamma_a^* \equiv \min \gamma_a^q, \text{ where } q \geq 1.$$

A.1 Infinite Horizon

To achieve infinite horizon objective T_c, the attacker only needs to enter c identities and can repeat the Sybil attack over multiple service phases without additional costs.

Whenever $q \geq c$ we have $\gamma_a^q = q$. The value for q that minimizes γ_a^q given $q \geq c$ is clearly c, and $\gamma_a^c = c$. So $\gamma_a^* = c$. This value does not involve n at all, so it is asymptotically c as n grows large.

A.2 Binomial

We consider Binomial objectives for three targets: a specific identity, any identity, or all identities. In each case the attacker is attempting to control at least c of k peers of the target identity or identities. Unfortunately, there is no closed-form expression for γ_a^* for the general c of k objective, which involves minimizing a complex expression involving generalized beta and polygamma functions. Instead, we find γ_a^* for the 1 of k and k of k objectives.

A.2.1 Binomial — Specific

1 **of** k First consider the 1 of k objective $B_{1,k}^{\text{spec.}}$. The probability of success given q identities is $1 - (\frac{n}{q+n})^k$, so

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

The q which minimizes γ_a^q must be either 1 or some root of the derivative of γ_a^q with respect to q. This derivative is given

$$\frac{\partial}{\partial q}\left(\frac{q}{1 - (\frac{n}{q+n})^k}\right) = \frac{n + q - (\frac{n}{n+q})^k (n + q + kq)}{(n+q)(-1 + (\frac{n}{n+q})^k)^2}$$

and is positive where

$$(n+q)((n+q)^k - n^k) - n^k kq > 0.$$

Assume that $k \geq 2$. Then $(n+q)^k \geq n^k + kqn^{k-1} + kq^{k-1}n + q^k$, and the left hand side of the preceding inequality is greater than or equal to

$$(n+q)(kqn^{k-1} + kq^{k-1}n + q^k) - n^k kq$$

and if this is greater than zero, the derivative must also be. Dividing both sides by $(n+q)$ we have

$$kqn^{k-1} + kq^{k-1}n + q^k - \frac{n^k}{n+q} kq > 0$$

and we know $\frac{n^k}{n+q} < n^{k-1}$ so the left hand side of the preceding inequality must be greater than

$$kqn^{k-1} + kq^{k-1}n + q^k - n^{k-1}kq$$

so the derivative is positive if

$$kq^{k-1}n + q^k > 0$$

which is always true for positive q and n. So when $k \geq 2$, the derivative is always positive, and the function γ_a^q is constantly increasing with increasing q. On the other hand if $k = 1$ then the derivative is positive where

$$(n+q)(n+q-n) - nq > 0$$

which simplifies to

$$\frac{q^2}{n+q} > 0$$

which is always true. So in this case as well the derivative is always positive, γ_a^q increases with increasing q.

Since γ_a^q is constantly increasing, the minimum must be at 1, the lowest possible value for q. So we have

$$\gamma_a^* = \gamma_a^1 = \frac{1}{1 - (\frac{n}{n+1})^k} \tag{19}$$

k of k We now consider the objective $B_{k,k}^{\text{spec.}}$. If $k = 1$ this is equivalent to $B_{1,k}^{\text{spec.}}$, and $\gamma_a^* = 1$. Hereafter we assume $k \geq 2$.

$$\gamma_a^q = \frac{q}{(\frac{q}{q+n})^k}$$

where the minimum value of q needed for success is k. The derivative is given

$$\frac{\partial}{\partial q}\left(\frac{q}{(\frac{q}{q+n})^k}\right) = (n - kn + q)\frac{(n+q)^{k-1}}{q^k}.$$

which has a root where $n - kn + q = 0$, at $q = (k-1)n$. We now confirm that $\gamma_a^{(k-1)n}$ is below γ_a^k, the other possible minima; we want to have

$$\begin{aligned}
\gamma_a^{(k-1)n} &\leq \gamma_a^k \\
\frac{k^k}{(k-1)^{k-1}}n &\leq k(\frac{k+n}{k})^k \\
k^{2k-1}n &\leq (k+n)^k(k-1)^{k-1} \\
(k^k)(\frac{k}{k-1})^{k-1}n &\leq (k+n)^k.
\end{aligned}$$

First note that if $k = 2$, the inequality holds so long as $n \geq 2$ (which is necessary in any case.) So in this case, $q = (k-1)n$ is the minimum q. Otherwise, assume $k \geq 3$. Note that $k^{k-1}/(k-1)^{k-1}$ is at most e, so the inequality holds if

$$k^k en \leq (k+n)^k.$$

The right hand side is at least

$$k^k + kk^{k-1}n + \frac{k(k-1)k^{k-2}{}^2}{n}$$

so the inequality holds if

$$\begin{aligned}
k^k en &\leq k^k + kk^{k-1}n + \frac{k(k-1)k^{k-2}{}^2}{n} \\
k^k en &\leq k^k + k^k n + \frac{k^k - k^{k-1}}{2}n^2 \\
en &\leq 1 + n + (\frac{1}{2} - \frac{1}{k})n^2.
\end{aligned}$$

The right hand side is at least

$$1 + n + (\frac{1}{2} - \frac{1}{3})n^2$$

so the inequality holds so long as

$$\begin{aligned} en &\leq \frac{n^2}{6} \\ n &\geq 17. \end{aligned}$$

So for reasonable n the inequality holds. (We do not consider the case $n < 17$.) We therefore have

$$\gamma_a^* = \gamma_a^{(k-1)n} = \left(\frac{k}{k-1}\right)^{k-1} kn \tag{20}$$

A.2.2 Binomial — Any

1 **of** k For the objective $B_{1,k}^{\text{any}}$ we have

$$\gamma_a^q = \frac{q}{1 - (\frac{n}{n+q})^{kn}}$$

and the derivative is given

$$\frac{\partial}{\partial q}\left(\frac{q}{1 - (\frac{n}{q+n})^{kn}}\right) = \frac{n + q - (\frac{n}{n+q})^{kn}(n + q + kq)}{(n+q)(-1 + (\frac{n}{n+q})^{kn})^2}.$$

The derivative is positive where

$$n + q - (\frac{n}{n+q})^{kn}(n + q + kq) > 0.$$

The left hand expression is always *greater* than the corresponding expression for the specific user case. So in this case as well, the derivative is always positive, γ_a^q increases with q, and the minimum value for γ_a^q occurs at $q = 1$; so

$$\gamma_a^* = \gamma_a^1 = \frac{1}{1 - (\frac{n}{n+1})^{kn}} \tag{21}$$

k **of** k For the objective $B_{k,k}^{\text{any}}$ we have

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

Unfortunately we have not been able to find the value of q that minimizes γ_a^q in this case. Through numerical experiment we have found a conjectural value of $q = (k-1)n + k$. At this value we have

$$\gamma_a^* = \frac{(k-1)n + k}{1 - (1 - (1 - \frac{n}{(n+1)k})^k)^n}. \tag{22}$$

A.2.3 Binomial – All

1 **of** k For the objective $B_{1,k}^{\text{all}}$ we have

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

The derivative is given

$$\frac{\partial}{\partial q}\left(\frac{q}{n(1 - (\frac{n}{q+n})^k)}\right) = \frac{n + q - (\frac{n}{n+q})^k(n + q + kq)}{n(n+q)(-1 + (\frac{n}{n+q})^k)^2}.$$

As the numerator is the same as in the specific case $B^{\text{spec.}}_{1,k}$ and the denominator is constantly positive, there are no roots in this case either. So the equilibrium valuation is at $q = 1$ and

$$\gamma_a^* = \gamma_a^1 \frac{1}{n(1 - (\frac{n}{n+1})^k)}. \tag{23}$$

k of k For the objective $B^{\text{all}}_{k,k}$ we have

$$\gamma_a^q = \frac{q}{n\frac{q}{q+n}k}.$$

The derivative is given

$$\frac{\partial}{\partial q}\left(\frac{q}{(\frac{q}{q+n})^k}\right) = (n - kn + q)\frac{(n+q)^{k-1}}{nq^k}.$$

Again the numerator is the same as in the specific victim case $B^{\text{spec.}}_{k,k}$ so the minimizing q must likewise be at $q = (k-1)n$. So we have

$$\gamma_a^* = \gamma_a^{(k-1)n} = \left(\frac{k}{k-1}\right)^{k-1} k. \tag{24}$$

A.3 Hypergeometric

As with the Binomial objective, it is not possible to give a closed-form expression for γ_a^* of $H_{c,k}$ with general c and k. There is no replacement when only a single choice is made, so $H_{1,k}$ is identical to $B^{\text{spec.}}_{1,k}$. For $H_{k,k}$ we have not been able to find the minimizing q, but based on numerical experiment we have a conjectural value of $q = (k-1)n + k$. At this value

$$\gamma_a^* = \frac{(nk-n)!(nk+k)!}{(nk)!((n+1)(k-1))!}. \tag{25}$$

A.4 Voting

In a Voting objective $V_{\alpha,\beta}$ the attacker attempts to control $\alpha(n+q)+\beta$ nodes total, but which particular nodes are controlled is not important. Solving for q we find that $q = \lceil \frac{\alpha n+\beta}{1-\alpha} \rceil$ is the minimum number of identities for success in the objective, and this number of identities guarantees success. So $\gamma_a^q = q$ and

$$\gamma_a^* = \lceil \frac{\alpha n + \beta}{1-\alpha} \rceil. \tag{26}$$