Sybilproof Reputation Mechanisms

Alice Cheng
Center for Applied Mathematics
Cornell University, Ithaca, NY 14853
alice@cam.cornell.edu

Eric Friedman
School of Operations Research and Industrial Engineering
Cornell University, Ithaca, NY 14853
friedman@orie.cornell.edu

ABSTRACT

Due to the open, anonymous nature of many P2P networks, new identities - or sybils - may be created cheaply and in large numbers. Given a reputation system, a peer may attempt to falsely raise its reputation by creating fake links between its sybils. Many existing reputation mechanisms are not resistant to these types of strategies.

Using a static graph formulation of reputation, we attempt to formalize the notion of sybilproofness. We show that there is no symmetric sybilproof reputation function. For nonsymmetric reputations, following the notion of reputation propagation along paths, we give a general asymmetric reputation function based on flow and give conditions for sybilproofness.

Categories and Subject Descriptors

C.2.4 [**Computer-Communication Networks**]: Distributed Systems

General Terms

Design, Theory

Keywords

sybils, reputation, peer-to-peer

1. INTRODUCTION

The large scale of many P2P networks and other online communities make it difficult for peers or users to assess the "trustworthiness" of other users, since a typical user's past history includes only a small fraction of the entire community. A common approach to this problem is the creation of a reputation system within the community. A reputation system attempts to aggregate the peers or users' collective experiences in order to allow a user to form an opinion about someone with whom he has not previously interacted [8]. Further, reputation may be used to assess and reward "good" behavior, for example, a peer or user with high reputation may be rewarded with preferred queueing or preferred partnering.

In recent years, many people have come up with reputation mechanisms for various applications [4, 5, 9, 7, 1], see also citations in [6]. However, much of this work has been ad-hoc - creating specific reputation mechanisms to combat some particular collection of adversarial strategies. In contrast, we would like to explore what conditions a reputation function must satisfy in order to be robust to adversarial strategies. This may allow a move towards defining a general class of robust reputation mechanisms. In this paper, we focus on the sybil attack.

In an online environment, new identities may be created with minimal cost, and users are not tied to unique identifiers. Therefore, a single user may create enough sybils to constitute a large fraction of the community [2]. In particular, a user may strategically create sybils for the purpose of boosting his own reputation. One example of this is the "link spamming" attack to PageRank [7] - when a single user attempts to boost his reputation by creating a large number of duplicate identities, who all recommend him.

Like many others [4, 9], we base our analysis on the trust network (or "web of trust") that may be constructed between users, where edges represent direct trust (i.e. the outcome of actual interactions). To aid analysis, we restrict our attention to reputation mechanisms that depend only the current structure of the trust graph. In section two, we summarize our main results. In section 3, we define such a reputation function on a trust graph, and define the notion of sybilproofness in this framework. In section 4, we discuss conditions for sybilproofness. In the case of symmetric reputation functions (i.e. functions invariant to graph isomorphism), we show that no sybilproof mechanism exists. In the asymmetric case, we give a generalized flow-based reputation function and give some conditions for sybilproofness in that case. In section 5, we conclude with some open questions in this area.

2. OVERVIEW OF RESULTS

While there has been much work related to reputation systems, many of these systems fail to be sybilproof. For example, in the well-known reputation system EigenTrust [5], users can typically increase their reputation values by creating complete subgraphs of sybils. Maximum flow from a fixed node s is another possible reputation function where a user may increase its relative rank via a sybil strategy. If we regard the edge values as capacities, then the maxi-

SIGCOMM'05 Workshops, August 22–26, 2005, Philadelphia, PA, USA.

mum flow from s to a node i is the maximum amount of flow we can push from s to i while respecting all capacity constraints. To see that max flow isn't sybilproof, notice that for any node i, several paths may contribute to i's reputation. Nodes that lie along those paths may have a worse reputation than i, and may be able to worsen the reputation of i by cutting off the path to i (for example, the node may split into two nodes, one taking all incoming links, one taking all outgoing links). Therefore, a node may be able to improve his rank with a sybil strategy. In this paper, we only consider sybil strategies where a user is only concerned with raising his own reputation. One might additionally consider "badmouthing" strategies where a user attempts to lower the reputation of others, without necessarily raising his own reputation. However, we conjecture that there is no reputation function which guards against all badmouthing strategies.

Related work includes [6], which has similar motivations.

Our primary interest in this paper is to provide conditions for which a reputation function is sybilproof. We show that symmetric reputation functions cannot be sybilproof, where a symmetric function is one which is invariant under a renaming of the nodes (i.e. it depends only on the edge structure of the graph). A sketch of the proof is as follows: If some node i doesn't have the highest possible reputation in a graph G, he can create a group of sybils, formed into an exact copy of G. (Note that we do not put restrictions on the number of sybils i may create, or on i's knowledge of the overall graph.) A symmetric function cannot distinguish between a node in the "original" graph and its corresponding node in the sybil copy, so for each node in the original part of the graph, there is some node in the sybil copy with the same reputation. Therefore, some sybil must have the highest reputation in the new graph. We can use a similar argument to show that symmetric reputation functions cannot be sybilproof even when we bound the number of allowable sybils by a constant. Intuitively, symmetric reputation functions fail to be sybilproof because a symmetric function cannot distinguish between groups of sybils and real nodes. In contrast, asymmetric reputation functions may assume that some specified nodes are trusted, and propagate trust from those nodes.

We also construct a formula for a class of asymmetric reputation functions. We build these functions from a real-valued function on paths, g, and a generalized addition operator $\oplus$. Like many existing reputation mechanisms [4, 9], this formulation employs the notion of transitive trust. This is the idea that if a user A trusts B, and B trusts C, then A may trust C to some extent, even if A hasn't previously interacted with C. Since direct trust is represented in the graph as directed edges, the notion of transitive trust exactly corresponds to propagation of trust along paths. Intuitively, g describes how trust propagates along a path, and $\oplus$ describes how to aggregate the trust among several paths. We show that when g and $\oplus$ satisfy certain properties, no user may increase his own reputation value with a sybil strategy. Furthermore, when we propagate trust along only a single "maximum" path, no user can improve his relative rank. Under some additional conditions for g, we show that f sybilproof implies that f propagates trust along a single path.

3. PROBLEM FORMULATION

We represent a network as a directed graph $G = (V, E)$, with users (some of whom may be sybils) represented as nodes, and interactions between a pair of users $i, j \in V$ represented as directed edges $e \in E$ between i and j with edge values $c(e) \in \mathbb{R}$ representing the outcomes of actual transactions. For this paper, we assume that $c(e)$ corresponds positively with positive transactions (so higher values correspond to higher trust). For example, $c(i, j)$ might be the number of times i successfully downloaded a file from j, or it may be a rating that i personally assigns to j. $\mathscr{G}$ is the collection of all such graphs, with a node set V a finite subset of $\mathbb{N} = \{0, 1, 2, \ldots\}$ and edge set E a subset of $V \times V$.

Given a graph $G = (V, E)$ we define the reputation to be a mapping on a graph associating each node with a real value.

Definition 1. A **reputation function** f is a real valued function on $\mathscr{G}$ that maps nodes of a graph to real values. We say that a node i in a graph G has reputation $f(G)_i \in \mathbb{R}$.

In a sybil strategy, a user is allowed to create arbitrarily many sybil nodes and fake edges between sybils. If U' is the collection of sybils, and $i \notin U'$, we require that any edge between i and U' represent real transactions, since i is not a member of the sybil group.

Definition 2. Given a graph $G = (V, E)$ and a user $i \in V$, we say that a graph $G' = (V', E')$ along with a subset $U' \subseteq V'$ is a **sybil strategy** for user i in the network $G = (V, E)$ if $i \in U'$ and that collapsing U' into a single node with label i in G' yields G. We can refer to U' as the sybils of i, and denote a sybil strategy by (G', U').

Note that in this formulation, if there is an edge $(j, i) \in E$ with value $c(j, i) = \alpha$, we allow in the sybil strategy of i the splitting of the edge to $(j, u_1), \ldots, (j, u_k)$ (with $u_1, \ldots, u_k$ sybils of i). However, we must split the value additively, so that $c(j, u_1) + \ldots + c(j, u_k) = \alpha$. Likewise, for any edge (i, j) with value α, we allow splitting of the edge to $(u_1, j), \ldots, (u_k, j)$ in a sybil strategy for i, such that $c(u_1, j) + \ldots + c(u_k, j) = \alpha$.

Reputations are often used to distinguish between users, and in many applications, users care more about relative reputation values than the actual values. Therefore, we say that a sybil strategy (G', U') for node i in G is successful, if some sybil has a better relative rank than i in G.

Since i may use any of his sybils to perform transactions in the future, a sybil strategy is successful if at least one sybil has higher rank. Thus, a reputation function is sybilproof if no successful sybil strategies are possible.

Definition 3. A reputation function f is **(rank) sybilproof** if for all graphs $G = (V, E)$, and all users $i \in V$, there is no sybil strategy for i, (G', U'), with $G' = (V', E')$ such that for some $u \in U'$, $\exists j \in V$ such that $f(G)_j > f(G)_i$ and $f(G')_u \geq f(G')_j$.

If f is rank sybilproof, then we will typically say that f is sybilproof. Note that in this definition, we allow a user to create arbitrarily many sybils. Though the cost of new identities is typically low in many settings, it is usually nonzero, so it is reasonable to consider cases where no user can create more than K sybils, for some K. This may allow a wider set of available reputation functions, as the set of K-sybilproof functions contain all sybilproof functions.

Definition 4. We say that a reputation function is **K-sybilproof** if it is sybilproof over all possible sybil strategies (G', U'), with $|U'| \leq K$.

Finally, in some settings, users may care more about their actual reputation values than their relative rank. For example, there are cases when users may earn specific rewards based on their reputation values. (e.g. Epinions paying users royalties based on their reputations [4]) To deal with these cases, we define the following:

Definition 5. A reputation function f is **value sybilproof** if for all graphs $G = (V, E)$ and all users $i \in V$ there is no sybil strategy for i, (G', U') such that for some $u \in U'$, $f(G)_i < f(G')_u$.

4. REPUTATION FUNCTIONS

In this section, we describe two types of reputation functions - symmetric and asymmetric - and give conditions for sybilproofness.

4.1 Symmetric Reputations

If anonymity is preserved within a reputation system for all users, it is sensible that under any renaming of the nodes, a reputation function should return the same values. More formally,

Definition 6. A reputation function f is **symmetric** if given a graph isomorphism σ, for all graphs $G = (V, E)$ with image G' under σ, and all nodes $i \in V$, $f(G)_i = f(G')_{\sigma(i)}$

In this treatment, the edge values encapsulate all information about actual interactions, so reputation should be determined only by the structure of the graph and the edge values. Therefore, two isomorphic graphs yield the same reputations.

Unfortunately, no symmetric reputation function is sybilproof, or even k-sybilproof for any $k \geq 2$. We may ignore the trivial reputation function which assigns to every node the same value, since it is trivially symmetric and sybilproof (but clearly unsatisfying as a reputation function candidate).

THEOREM 1. *There is no symmetric sybilproof nontrivial reputation function*

PROOF. Given a graph $G = (V, E)$, and a nontrivial global reputation function f, let $i \in V$ be a node which doesn't have the highest reputation in V. f is nontrivial, so we can find a G such that such a node exists. Note that for any f, G, since V is finite, there always exists a node $i \in V$ that attains the maximum value of $f(G)_i$ among nodes in V. Consider the sybil strategy of i, (G', U') such that $G \subseteq G'$ and for all nodes j, there is an additional copy of j, $j' \in V'$, such that for all edges $(j, k) \in E$ with edge value $c(j, k)$, there is an edge $(j', k') \in E'$ with value $c(j, k)$. In other words, G' is a disjoint union of two copies of G. If $U' = i \cup \{j' \mid j \in V\}$, then by symmetry, there is some $u \in U'$ that attains the maximal reputation: $f(G')_u \geq f(G')_j$ for all $j \in V$. □

Not only is a symmetric reputation function not sybilproof, the above proof shows that if G is such that not all nodes have the same reputation value, any node that doesn't have the highest possible value has a successful sybil strategy.

To show the impossibility of k-sybilproofness, we assume a k-sybilproof function exists, and rather than creating the duplicate graph in one step as in the above proof, we create it over many steps, applying k-sybilproofness at each step. We then get a contradiction at the final step by applying symmetry.

THEOREM 2. *There is no nontrivial symmetric k-sybilproof reputation function for any constant $k > 1$.*

PROOF. It suffices to show the result for $k = 2$. For a contradiction, suppose a symmetric 2-sybilproof nontrivial reputation function f exists. Let $G = (V, E)$ be a network, where there is some user $u \in V$ such that there exists a $v \in V$ with $f(G)_v > f(G)_u$. Let $n = |V|$, and WLOG, let $V = \{0, \ldots, n-1\}$. There exist graphs $G_1 = (V_1, E_1), \ldots, G_n = G = (V, E)$, where G_{i-1} is constructed from G_i by contracting nodes $i-2$ and $i-1$ and relabelling it $i-2$. We can relabel the vertices in G_i so that $V_i = \{n, \ldots, n+i-1\}$ so that vertex j is relabelled as $n+j$, and let $G_i^* = G \cup G_i$. From the proof of Theorem 1, there is a node $v \in G_n$ such that $f(G_n^*)_v$ is maximal over V_n. We can relabel the nodes so that $v = 2n-1$.

By construction, G_i^* is a sybil strategy in G_{i-1}^* for node $n+i-2$ for each $i = 1, \ldots, n$, and G_1^* is a sybil strategy in G for node u. Since f is sybilproof, node $n \in G_1$ cannot have the highest reputation in V_1, and by induction, for each i, node $n+i-1$ cannot have the highest reputation in V_i. Therefore, $f(G_n^*)_v$ is not maximal over V_n, a contradiction. Thus, if f is a symmetric nontrivial reputation function, it cannot be 2-sybilproof. □

A parallel argument shows an similar result for value sybilproof reputation functions. However, we need to exclude certain pathological reputation functions. One simple criteria which is satisfied by all nontrivial reputation functions is based on the following construction: Given a graph G define its B-extension with respect to i, $B_i(G)$ to be the graph which is constructed by taking a copy of G and contracting the node $i \in V$ with its double in the copy of G.

Definition 7. A reputation function f is **B-Nontrivial** if there exists a graph $G = (V, E)$ and $i, j \in G$ such that $f(G)_j > f(G)_i$ and $\exists v \in V'$ such that $f(G')_v > f(G')_i$, where G' is the B-extension of B with respect to i.

THEOREM 3. *If a reputation function f is B-nontrivial then it cannot be value sybilproof, or k-value sybilproof.*

Note that PageRank is B-nontrivial, so this shows that symmetric variants of PageRank are not sybilproof or value sybilproof.

4.2 Asymmetric Reputations

To construct an asymmetric sybilproof reputation function, note that we can easily break symmetry by computing reputation values with respect to some fixed node in the graph. This may be useful when we can identify some trusted user, or when each user computes separately the reputations of other users with respect to themselves.

Once we have a fixed root node s, we may allow reputation to propagate along paths outward from s. Similar ideas also motivated the flow-based reputation systems described in [3, 9]. Given a fixed graph $G = (V, E)$ and a root node $s \in V$, let $\mathbb{P}_i$ be the set of all collections of edge-disjoint paths from

s to i in G. We allow an edge of value $\alpha+\beta$ to split into two parallel edges with values α, β at will. Let g be a function from paths to real numbers, and let $\oplus$ be an "addition"-like operator on real numbers.

Given root node s, $G = (V, E)$ and $i \in V$, $i \neq s$, define

$$(f^s(G))_i := \max_{\mathscr{P}_{s,i} \in \mathbb{P}_i} \bigoplus_{P \in \mathscr{P}_{s,i}} g(P)$$

We set $f^s(G)_s = \infty$.

Reputation propagates along disjoint paths in some manner (described by the function g), and then the reputation values are aggregated by $\oplus$. If we regard edge values as edge capacities, flow-based reputation functions may be considered. For example, one can show that maximum flow falls under this category of functions, if we let $g(P) = \min\{c(e) | e \in P\}$ and $\oplus = +$. Many flow variants are included in this class, such as maximum generalized flow (the case where flow leaks along an edge), or the variation on max flow preferring closer nodes used in [3]. Other possible functions in this framework include the maximum capacity path and $1/\#hops$ for the minimum hop path.

Given some restrictions on $g, \oplus$, we can show that f^s is value sybilproof:

THEOREM 4. *If f^s as defined above satisfies the following properties,*

(a) (Diminishing returns) For all $s-i$ paths P, if an $s-j$ path P' is an extension of P, then $g(P') \leq g(P)$.

(b) (Monotonicity) $\oplus$ is nondecreasing, and g is nondecreasing with respect to the edge values.

(c) (No splitting) Given a single $s-i$ path P, if we split P into two $s-i$ paths P_1, P_2, then $g(P_1) \oplus g(P_2) \leq g(P)$.

for all graphs $G = (V, E)$, $s \in V$, and all $i \in V$, then f^s is value sybilproof.

PROOF. Let $G = (V, E)$ be a graph, let $s \in V$, and $i \in V, i \neq s$. Let (G', U') be a sybil strategy for i with respect to f^s, with collection of sybils U'. For $u \in U'$, there is some collection of disjoint $s-u$ paths $\mathscr{P}$ in G' such that $f^s(G')_u = \bigoplus_{P \in \mathscr{P}} g(P)$. For each $P \in \mathscr{P}$, let P' be the subpath starting from s and ending at the first node in U' appearing along the path. By (a), $g(P') \geq g(P)$, and by the definition of a sybil strategy, P' must correspond to some $s-i$ path in G. Let $\mathscr{P}' = \{P' \mid P \in \mathscr{P}\}$. $\mathscr{P}'$ forms an edge disjoint collection of $s-i$ paths in G. So, by the definition of f^s,

$$f^s(G)_i \geq \bigoplus_{P' \in \mathscr{P}'} g(P') \geq \bigoplus_{P \in \mathscr{P}} g(P) = f^s(G')_u$$

If $i = s$, $f^s(G)_i = \infty$, so $f^s(G')_u$ cannot have a higher value for any sybil strategy (G', U'), and $u \in U'$. □

For f^s satisfying the above conditions, in any graph G with base node s, no node i can increase his reputation value, $f^s(G)_i$. However, a node may still be able to improve his rank if he can sufficiently lower the reputation values for certain other nodes. A node i may not improve by this method if the only nodes who may be affected by i's sybil strategies have lower reputation than i, since lowering their reputation doesn't affect i's rank. Thus, propagating reputation only along a maximum g-valued single path gives a sybilproof function. $\oplus = \max$ gives this single path reputation function, so we have the following:

THEOREM 5. *If f^s satisfies the above properties and additionally, $\oplus = \max$, then f^s is sybilproof. Conversely, if g is such that for all paths P, there exists a strictly longer path P', $P \subset P'$, such that $g(P) = g(P')$, then f^s sybilproof implies that $\oplus = \max$.*

PROOF. By theorem 4, there is no sybil strategy to increase one's own reputation value, so it suffices to show that there is no sybil strategy to reduce the reputation value of higher ranked users. By the definition of a sybil strategy, and the monotonicity of $\oplus$, at worst, a node i can affect a node j by removing all the $s-j$ paths that pass through i. However, i may only affect j's reputation value if i lies along the $s-j$ path P with maximum value $g(P)$ over all $s-j$ paths in G. If P' is an $s-i$ path contained in P, $f^s(G)_i \geq g(P')$, but $f^s(G)_j \leq g(P')$, so $f^s(G)_j \leq f^s(G)_i$, and i cannot increase his rank by by lowering j's reputation value.

For the other direction, suppose $\oplus > \max$. Let paths $P_1, \ldots, P_k$ be the minimal set of paths such that $g(P_1) \oplus \ldots \oplus g(P_k) > \max(g(P_1), \ldots, g(P_k))$. Note that $k > 1$, since $g(P_1) = \max(g(P_1))$. For convenience, let $g(P_1) \geq g(P_2) \ldots \geq g(P_k)$, so

$$v = g(P_1) \oplus \ldots \oplus g(P_k) > g(P_1)$$

Since this set is minimal,

$$g(P_2) \oplus \ldots \oplus g(P_k) = \max(g(P_2), \ldots, g(P_k)) \leq g(P_1) < v$$

By assumption on g, there exists a strictly longer path P_1' such that P_1 is a subpath of P_1' and $g(P_1') = g(P_1)$. Let G be the graph such that s and i have k parallel $s-i$ paths: $P_1', P_2, \ldots, P_k$. There is some node j on P_1' such that the only $s-j$ path is P_1. By no splitting rule, $f^s(G)_j = g(P_1)$. $f^s(G)_i \geq g(P_1') \oplus \ldots \oplus g(P_k) > g(P_1)$. Consider the sybil strategy for j where j splits into two nodes, one accepting all incoming arcs to j, and the other accepting all outgoing arcs from j. In this new graph G', i is only connected to s via $P_2, \ldots, P_k$, so $f^s(G')_i = g(P_2) \oplus \ldots \oplus g(P_k) = g(P_2) \leq g(P_1)$. The sybil node for j accepting all incoming arcs clearly has the same reputation as j, and for all other nodes, their reputation may not increase. Therefore, this sybil strategy gives j a strictly higher rank, so f^s is not sybilproof. Therefore, $\oplus \leq \max$.

Suppose $\oplus < \max$. Then, for some collection of disjoint paths $P_1, \ldots, P_k$ with $g(P_1) \geq \ldots \geq g(P_k)$, $g(P_1) \oplus \ldots \oplus g(P_k) < g(P_1)$. Therefore, when we maximize over all collections of disjoint paths, we only choose collections with only one path, so replacing $\oplus$ with max gives the same function. Therefore, $\oplus = \max$. □

Note that we cannot entirely remove the requirement on g in Theorem 5. For example, consider the case where $c(e) = 1$ for all e, $g(P) = \frac{1}{\text{hop length of } P}$ if P is composed only of edges with capacity 1, and $v_1 \oplus \ldots \oplus v_k = \max(v_1, \ldots, v_k) + \epsilon(\#v_i$ that attain the maximum value), for ϵ some infinitesimal value. One can check that this yields a sybilproof f^s, and clearly, $\oplus \neq \max$.

5. CONCLUSIONS

We have presented a possible framework for assessing a reputation mechanism's robustness to sybils. We have shown that no nonconstant, symmetric reputation function exists.

Further, we have given a collection of flow-based asymmetric reputation functions which are sybilproof, under some conditions.

There are many open questions related to this direction of work. For one, the flow-based reputation function defined on section 4.2 is unlikely to be the most general formulation of an asymmetric reputation function. Finding a more general formulation, along with necessary and sufficient conditions for sybilproofness would be useful. Furthermore, in this paper, we have focused strictly on a static graph model of reputation. One possible generalization of this framework is to allow reputation functions to depend on the state of the network at previous time steps as well as the current state of the network. It would also be interesting to extend the analysis here to explore necessary requirements for sybilproofness in a dynamic reputation model.

6. REFERENCES

[1] S. Buchegger and J. Le Boudec. A robust reputation system for p2p and mobile ad-hoc networks. In *Second Workshop on the Economics of Peer-to-Peer Systems*, 2004.

[2] J. Douceur. The sybil attack. In *Proc. of the IPTPS02 Workshop*, 2002.

[3] M. Feldman, K. Lai, I. Stoica, and J. Chuang. Robust incentive techniques for peer-to-peer networks. In *ACM E-Commerce Conference (EC'04)*, 2004.

[4] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th International World Wide Web Conference*, pages 403–411, 2004.

[5] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th International World Wide Web Conference*, pages 640–651, 2003.

[6] R. Morselli, J. Katz, and B. Bhattacharjee. A game-theoretic framework for analyzing trust-inference protocols. In *Second Workshop on the Economics of Peer-to-Peer Systems*, 2004.

[7] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. In *Proceedings of the 7th International World Wide Web Conference*, pages 161–172, 1998.

[8] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation systems. In *Communications of the ACM*, volume 43, pages 45–48, 2000.

[9] M. Richardson, R. Agrawal, and P. Domingos. Trust management for the semantic web. In *Proceedings of the Second International Semantic Web Conference*, pages 351–368, 2003.