On the Sybil-Proofness of Accounting Mechanisms

Sven Seuken
School of Engineering & Applied Sciences
Harvard University
33 Oxford St., Cambridge, MA
seuken@eecs.harvard.edu

David C. Parkes
School of Engineering & Applied Sciences
Harvard University
33 Oxford St., Cambridge, MA
parkes@eecs.harvard.edu

ABSTRACT

A common challenge in distributed work systems like P2P file-sharing communities, or ad-hoc routing networks, is to minimize the number of free-riders and incentivize contributions. Without any centralized monitoring it is difficult to distinguish contributors from free-riders. One way to address this problem is via accounting mechanisms which rely on voluntary reports by individual agents and compute a score for each agent in the network. In Seuken et al. [11], we have recently proposed a mechanism which removes any incentive for a user to manipulate the mechanism via misreports. However, we left the existence of *sybil-proof* accounting mechanisms as an open question. In this paper, we settle this question, and show the striking impossibility result that under reasonable assumptions no sybil-proof accounting mechanism exists. We show, that a significantly weaker form of K-sybil-proofness can be achieved against certain classes of sybil attacks. Finally, we explain how limited robustness to sybil manipulations can be achieved by using max-flow algorithms in accounting mechanism design.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics*

General Terms

Algorithms, Design, Economics

Keywords

Mechanism Design, Sybil-Proofness, P2P, Reputation

1. INTRODUCTION

Distributed work systems arise in many places, for example in peer-to-peer (P2P) file-sharing networks, or in ad-hoc wireless networks where individual peers route data packages for each other. Of course, the total amount of work performed by a population is equal to the total amount of work consumed. Moreover, while some degree of free-riding may be acceptable, the long-term viability of distributed work systems relies on roughly balanced work contributions. Otherwise, strategic agents may seek to free-ride on the system, i.e., minimize the work they perform and maximize the work they consume. This problem becomes particularly challenging when the interactions are bilateral, there is no a priori trust relation between the agents, there is no ability to monitor activities, no contract covers the interactions, and no currency can be used because the institutional requirements for payment exchanges are not available.

In Seuken et al. [11], we have recently formalized this problem and introduced *accounting mechanisms* that provide a solution: they keep long-term tallies of work performed and consumed and compute a score that approximates an agent's net contributions. Because accounting mechanisms rely on voluntary reports, a major challenge is to provide robustness against strategic manipulations. The two manipulations we consider are *misreports*, where an agent overstates the amount of work contributed or consumed, and *sybil manipulations*, where an agent creates *fake* copies of itself. Previously, we have proposed the *Drop-Edge* mechanism which selectively ignores some of the reports, thereby providing misreport-proofness [11]. However, we have left open the question whether *sybilproof* accounting mechanism exist. In this paper, we prove that under reasonable assumptions, no sybilproof accounting mechanism exists. We show that a significantly weaker form of robustness can be achieved for a restricted class of attacks. Finally, we discuss the usefulness of max-flow algorithms for limited robustness against sybil manipulations in practice.

Accounting vs. Reputation Mechanisms.
Misreport and sybil manipulations are well-studied in the related literature on trust and reputation mechanisms [6]. However, the results from this literature do not translate to accounting mechanisms. First, in distributed work systems, every *positive* report by A about his interaction with B, i.e., B performed work for A, is simultaneously a *negative* report about A, i.e., A received work from B. This fundamental tension is not present in reputation mechanisms. Second, sybil manipulations are much more powerful against accounting mechanisms. For a search engine, for example, the primary concern is that an agent could increase the reputation of its website by creating a set of sybils that are linking to the original website, but an agent does not care about the reputation of the sybils themselves. In a distributed work system, in contrast, if an agent can create sybils with a positive score, then these sybils may receive work from other users without negatively affecting

NetEcon'11, June 5, 2011, San Jose, California, USA.

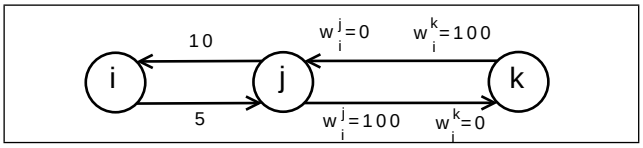

Figure 1: A subjective work graph from agent i's perspective. Edges where i has direct information have only one weight. Other edges can have two weights, corresponding to the possibly conflicting reports of the two agents involved.

the score of the original agent. While various reputation mechanisms have been proposed that are sybil-proof (e.g., maxflow, hitting-time [2, 12]), these results do not translate to accounting mechanisms. Third, once an agent has a high reputation it can benefit from that for a long time. For example, a website with a high PageRank [8] benefits from lots of visitors without affecting its reputation. In distributed work systems, in contrast, an agent benefits from a high score by getting priority for receiving work in the future, which in turn decreases its score again. Thus, accounting scores are inherently temporary.

Related Work.
Despite the differences between accounting and reputation mechanisms, the literature on transitive trust and reputation mechanisms [6] is an important precursor to our own work. One of the largest steps forward regarding robust incentives in a real-world P2P system was the BitTorrent protocol [3]. In contrast to previous protocols like Napster or Gnutella, BitTorrent uses a policy with short-term, direct incentives, resembling to a large degree a simple tit-for-tat mechanism. Feldman et al. [4, 5] study the challenges involved in providing robust incentives in fully decentralized P2P networks. However, they do not propose a misreport-proof mechanism. Piatek et al. [9] find empirically that most users of P2P file-sharing networks are connected via a one hop link in the connection graph and motivate the use of well-connected intermediaries to broker information. Along similar lines, Meulpolder et al. [7] present a fully decentralized accounting mechanism, but without a formal analysis of its properties. Cheng et al. [1, 2] have studied the sybil-proofness of reputation mechanisms. While their work influenced our thinking about sybil-proofness, unfortunately, their results do not translate to our domain, due to the differences between accounting and reputation mechanisms. Resnick and Sami [10] also study the sybil-proofness problem. However, in their model, individual transactions are risky and can have positive or negative outcomes. In contrast, in our domain, individual transactions are not risky. Instead, our focus is on computing accounting scores that are proportional to the net work contributed by the agents.

2. FORMAL MODEL

Consider a distributed work system of n agents each capable of doing work for each other. All work is assumed to be quantifiable in the same units. The work performed by all agents is captured by a work graph:

Definition 1. **(Work Graph)** A work graph $G = (V, E, w)$ has vertices $V = \{1, \ldots, n\}$, one for each agent, and directed edges $(i, j) \in E$, for $i, j \in V$, corresponding to work performed by i for j, with weight $w(i, j) \in \mathbb{R}_{\geq 0}$ denoting the number of units of work.

Figure 2: Accounting Mechanism and Allocation Policy.

We use $e \in E$ when referring to a generic edge, and $(i, j) \in E$ when referring to the specific edge from i to j. The true work graph may be unknown to individual agents who only have direct information about their own participation:

Definition 2. **(Agent Information)** Each agent $i \in V$ keeps a private history $(w_i(i, j), w_i(j, i))$ of its interactions with other agents $j \in V$, where $w_i(i, j)$ and $w_i(j, i)$ are the work performed for j and received from j respectively.

Based on its own experiences and known reports from other agents, agent i can construct a subjective work graph (see Figure 1). Let $w_i^j(j, k), w_i^k(j, k) \in \mathbb{R}_{\geq 0}$ denote the edge weight as reported by agent j and agent k respectively.

Definition 3. **(Subjective Work Graph)** A subjective work graph from agent i's perspective, $G_i = (V_i, E_i, w_i)$, is a set of vertices $V_i \subseteq V$ and directed edges E_i. Each edge $(j, k) \in E_i$ for which $i \notin \{j, k\}$, is labeled with one, or both, of weights $w_i^j(j, k), w_i^k(j, k)$ as known to i. For edges (i, j) and (j, i) the associated weight is $w_i^i(i, j) = w(i, j)$ and $w_i^i(j, i) = w(j, i)$ respectively.

We assume that these weights are shared through voluntary reports to a central server, while still maintaining the core assumption of no central monitoring and no independent verification of reports. Thus, the edge weights $w_i^j(j, k)$ and $w_i^k(j, k)$ need not be truthful reports about $w(j, k)$. In earlier work we also considered a decentralized protocol, with bilateral sharing of information with other agents [11].

Periodically, an agent can receive a work request by a set of agents with which the agent may have rarely or never interacted with before. This induces a choice set:

Definition 4. **(Choice Set)** We let $C_i \subseteq V \setminus \{i\}$ denote the choice set for agent i, i.e., the set of agents that are currently interested in receiving some work from i.

An accounting mechanism computes a *score* for each agent $j \in C_i$, proportional to the net work contributed.

Definition 5. **(Accounting Mechanism)** An accounting mechanism M takes as input a subjective work graph G_i, a choice set C_i, and determines the score $S_j^M(G_i, C_i) \in \mathbb{R}$, for any agent $j \in C_i$, as viewed by agent i.

We let S_0^M denote the *default* score that accounting mechanism M assigns to an agent about which no information regarding work consumed or performed is available (i.e., the two agents are disconnected in the subjective work graph). Once the accounting mechanism has computed a score for each agent in the choice set, an agent needs an *allocation policy* (e.g., "winner-takes-all" or "threshold rules") to decide to whom to allocate work to (see Figure 2).

Definition 6. **(Allocation Policy)** Given a choice set C_i and accounting scores S_j^M for each agent $j \in C_i$, an allocation policy $A(C_i, S^M)$ selects one agent $j^* \in C_i$ for whom agent i shall perform one unit of work.

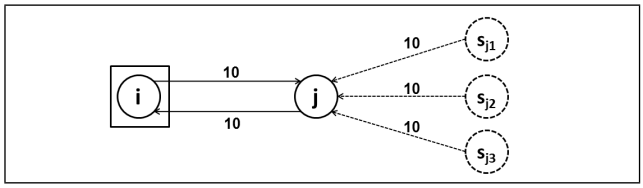

Figure 3: A strongly beneficial sybil attack.

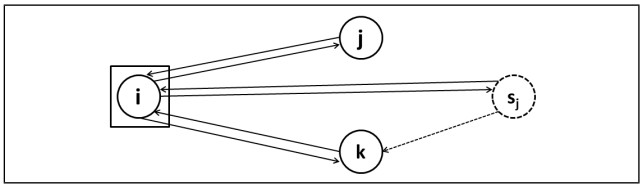

Figure 4: A weakly beneficial sybil attack.

3. STRATEGIC MANIPULATIONS

We assume an agent population that consists of *cooperative* agents who contribute approximately as much as they consume work, and *free-riders* who contribute less than they consume. Accounting mechanisms help agents differentiate between cooperative agents and free-riders, in such a way that the performance increases for cooperative agents and decreases for free-riders. For a well functioning accounting mechanism we need to remove advantages from strategic manipulations. The first class of manipulations we consider are *misreports* [11], where an agent reports false information about its work performed or consumed:

Definition 7. **(Misreport-proof)** An accounting mechanism M is misreport-proof if, for any subjective work graph G_i, any choice set C_i, any agent $j \in C_i$, for every misreport manipulation by j, where G_i' is the subjective work graph induced by the misreports, the following holds:

- $S_j^M(G_i', C_i) \leq S_j^M(G_i, C_i)$, and
- $S_k^M(G_i', C_i) \geq S_k^M(G_i, C_i)\ \forall k \in C_i \setminus \{j\}$.

In words, we consider a mechanism to be misreport-proof if reporting false work information can only worsen an agent's own score and improve the score of other agents.

The second class of manipulations we consider are *sybil manipulations*, where an agent introduces sybils (fake agents) into the network to manipulate the accounting mechanism. Given subjective work graph G_i, an attacking agent can do multiple things, e.g., add sybils to the network, or make multiple false reports. We model this as happening in one step, inducing a new subjective work graph G_i'.

Definition 8. **(Passive Sybil Attack)** A passive sybil attack by agent j is a tuple $\sigma_j = (V_s, E_s, w_s)$ where $V_s = \{s_{j_1}, s_{j_2}, ...\}$ is a set of sybils, $E_s = \{(x, y) : x, y \in S \cup \{j\}\}$, and w_s are the edge weights for the edges in E_s (one weight per edge). Applying the sybil attack σ_j to agent i's subjective work graph $G_i = (V_i, E_i, w_i)$ results in a modified subjective work graph $G_i \downarrow \sigma_j = G_i' = (V_i \cup V_s, E_i \cup E_s, w')$ where $w'(e) = w_i(e)$ for $e \in E_i$ and $w'(e) = w_s(e)$ for $e \in E_s$.

This attack is called *passive*, because the sybils themselves do not perform work. Attacks where some of the sybils also perform work, possibly over multiple time steps, are called **active sybil attacks**. The following definitions are purposefully written to be agnostic to the behavior of other agents in the network. Our theoretical results hold without requiring specific assumptions about these behaviors. In particular, the negative results hold even assuming that all other agents behave truthfully, and the positive result holds even for arbitrary behavior of other agents.

We consider a sybil attack to be *beneficial* if as a result of the manipulation, the attacking agent or one of its sybils is selected to receive some work when it previously didn't:

Definition 9. **(Beneficial Sybil Attack)** Given accounting mechanism M, subjective work graph G_i, and choice set C_i such that agent i's allocation policy picks agent $k \in C_i$, i.e., $A(G_i, C_i) = k$, a beneficial (passive or active) sybil attack σ_j (with at least one sybil s) by agent $j \neq k \in V_i$ such that $G_i \downarrow \sigma_j = G_i'$ and $C_i \downarrow \sigma_j = C_i'$, is one where at least one of (1), (2), (3), or (4), or combinations thereof holds:

(1) j's score is increased such that now $A(G_i', C_i') = j$.

(2) other agents' scores are lowered such that $A(G_i', C_i') = j$.

(3) sybil s is created with a score such that $A(G_i', C_i') = s$.

(4) other agents' scores are lowered such that $A(G_i', C_i') = s$.

How beneficial an attack really is depends on the trade-off between the amount of work necessary to perform the attack, and the resulting amount of "free" work to be consumed:

Definition 10. **(Strongly vs. Weakly Beneficial Sybil Attacks)** Given accounting mechanism M and work graph $G_i = (V_i, E_i, w_i)$, assume agent $j \in V_i$ performs a (passive or active) sybil attack σ_j such that $G_i' = \sigma_j(G_i)$. Let σ_j^n denote an $n-$times-repetition of the sybil attack. Let $\omega^-(\sigma_j^n)$ denote the amount of work involved in performing σ_j^n, and let $\omega^+(\sigma_j^n)$ denote the amount of work that agent j or any of its sybils will be able to consume. We call σ_j a:

- **strongly beneficial sybil attack:** if $\omega^+(\sigma_j^n) > 0$, and $\omega^-(\sigma_j^n) = 0$ or $\lim_{n\to\infty} \frac{\omega^+(\sigma_j^n)}{\omega^-(\sigma_j^n)} = \infty$.
- **weakly beneficial sybil attack:** if $\omega^+(\sigma_j^n) > 0$ and $\omega^-(\sigma_j^n) > 0$, and $\exists c \in \mathbb{R}_{\geq 0} : \lim_{n\to\infty} \frac{\omega^+(\sigma_j^n)}{\omega^-(\sigma_j^n)} \leq c$.

Of course, whether a particular sybil attack is beneficial or not, depends on the accounting mechanism used. Yet, in Figures 3 and 4 we give two generic examples. In Figure 3, agent j has already performed/consumed 10 units of work for/from agent i, and we assume that i now trusts j's reports about other agents to some degree. Now, j creates a set of sybils and falsely reports to i that these sybils have performed 10 units of work for j. Assuming that this raises the sybils' scores high enough (property (3) of Definition 9), each sybil can now exploit its score and consume some work from i. Once the sybils' scores are "used up", j can simply create another sybil and repeat the attack, without performing new work. Thus, this constitutes a strongly beneficial (passive) sybil attack.

In Figure 4, we assume that j and k are both inside i's choice set. Now, j performs an active sybil attack, creating sybil s_j and letting s_j perform some work for i and consume some work for i. Next, s_j makes a false report to i, claiming that agent k has consumed a lot of work from s_j. This may now decrease agent k's score enough, such that j is chosen by the allocation policy the next time j and k compete for

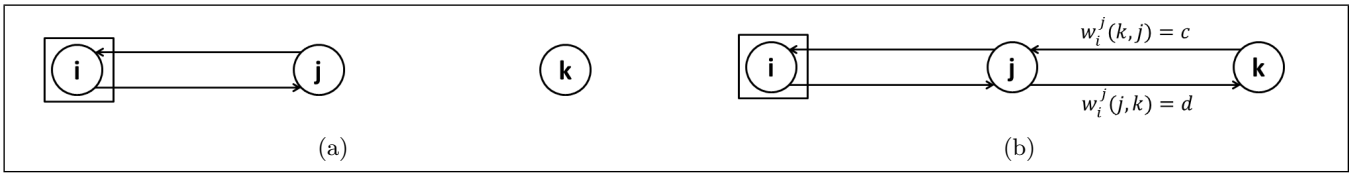

Figure 5: An illustration of single-report-responsiveness: there exists a subjective work graph G_i, e.g., the one shown in (a), such that a single positive report by j about k, as shown in (b), leads to $S_{ik}^M(G_i, C_i) > S_0^M$.

some of i's work (property (2) of Definition 9). However, consuming work decreases j's score, which means that after a certain amount of consumption, i's allocation policy will pick k instead of j. Thus, this attack constitutes a *weakly* beneficial active sybil attack. Note that in general, passive and attacks may be weakly or strongly beneficial.

Definition 11. **(Sybil-Proofness)** An accounting mechanism is sybil-proof against strongly (weakly) beneficial sybil attacks, if for every work graph G_i there exists no passive or active strongly (weakly) beneficial sybil attack.

We now introduce three natural axioms regarding the design of accounting mechanisms. First:

Definition 12. **(Independence of Disconnected Agents)** Accounting mechanism M satisfies independence of disconnected agents, if for any subjective work graph $G_i = (V_i, E_i, w_i)$ and any choice set C_i, for any $k \in V_i$ for which there does not exist an edge in E_i or for which all edges in E_i have zero weight, where G_i' denotes the graph where node k has been removed, the following holds:

$$\forall j \in V_i' : S_{ij}^M(G_i, C_i) = S_{ij}^M(G_i', C_i')$$

This axiom requires that the scores do not depend on disconnected agents, i.e., adding or removing agents with no amount of work consumed or performed does not change the scores of other agents.

Definition 13. **(Symmetric Accounting Mechanisms)** An accounting mechanism M is symmetric if for any subjective work graph $G_i = (V_i, E_i, w_i)$ and choice set C_i, any graph isomorphism f such that $G_i^{'} = f(G_i)$, $C_i^{'} = f(C_i)$ and $f(i) = i$:

$$\forall j \in V_i \setminus \{i\} : S_{ij}^M(G_i, C_i) = S_{if(j)}^M(G_i^{'}, C_i^{'}).$$

In words, this axiom requires that a priori, the accounting mechanism does not put more or less trust into any agent, i.e., we only consider mechanisms that, for any renaming of the agents in the network, return the same scores.[1]

Our third axiom excludes any "trivial" accounting mechanisms that assign the same or random scores to every agent, as well as mechanisms that ignore all information except an agent's own direct experiences.

[1]In the context of reputation mechanisms [1], *symmetry* typically corresponds to globally consistent, or *objective* reputation values, where every agent in a network has the same view on each other agent's reputation, in contrast to *asymmetric* mechanisms that allow for *subjective* reputation values. In this terminology, our accounting mechanisms are all "asymmetric" because they all inherently lead to subjective scores as they are based on subjective work graphs. However, what we mean by "symmetry" is something different, namely that from each individual agent's perspective, other agents' scores are invariant to identities.

Definition 14. **(Single-Report Responsiveness Property)** An accounting mechanism M has the single-report responsiveness property if, for any agent i, there exists a subjective work graph $G_i = (V_i, E_i, w_i)$ and choice set C_i, with nodes j and k such that nodes i and j are neighbors in G_i and no path connects nodes i and k, and there exists a graph $G_i' = (V_i', E_i', w_i')$ with $V_i' = V_i$, $E_i' = E_i \cup \{(k,j),(j,k)\}$, and $w_i'(e) = w_i(e)$ for all $e \in E_i \setminus \{(k,j),(j,k)\}$, and there exists a constant $c \in \mathbb{R}_{>0}$ with $w_i'^j(k,j) = c$, such that:

$$S_k^M(G_i', C_i') > S_0^M.$$

In words, this axiom requires that there exists a work graph where i has no information about k, and where a *single* positive report by agent j about agent k can increase the score that i assigns to k above the default score (see Figure 5).

3.1 Impossibility of Sybil-Proofness

THEOREM 1. *For every accounting mechanism that satisfies independence of disconnected agents, is symmetric, has the single-report responsiveness property, and is misreport-proof, there exists a passive strongly beneficial sybil attack.*

PROOF. Assume that accounting mechanism M satisfies the single-report responsiveness property. Thus, there exists a graph G_i and nodes i, j and k as described in Definition 14, for example like the one depicted in Figure 6 (a), and in particular $A(G_i, C_i) \neq k$. Now, let agent j create a sybil node s_j and insert it into G_i such that $G_i' = (V_i', E_i, w_i)$ with $V_i' = V_i \cup \{s_j\}$. Because of the independence of disconnected agents, the scores of all agents in the graph have remained the same. Note that there is no path connecting k and i as well as no path connecting s_j and i, and thus the two nodes k and s_j look the same from i's perspective. Now, assume that agent k performs c units of work for j (as needed for the single-report responsiveness property), and by misreport-proofness, agent j is best off making a truthful report to i about this interaction, leading to subjective work graph G_i'' such that $S_{ik}^M(G_i'', C_i) > S_0^M$. We assume that $S_{ik}^M(G_i'', C_i)$ is large enough such that now $A(G_i'', C_i) = k$. Because M is symmetric, we can apply a graph isomorphism f to G_i'' that only switches the labeling of s_j and k. Thus, there exists a report that j can make about s_j with $w_i'^j(s_j, j) = c$ leading to graph G_i^* such that $S_{is_j}^M(G_i^*, C_i) > S_0^M$ (see Figure 6 (b)). Because of misreport-proofness, we know that agent j has no disadvantage from making this report. Thus, property (3) of Definition 9 is satisfied, and because the attack itself involves no work, this constitutes a strongly beneficial sybil attack. □

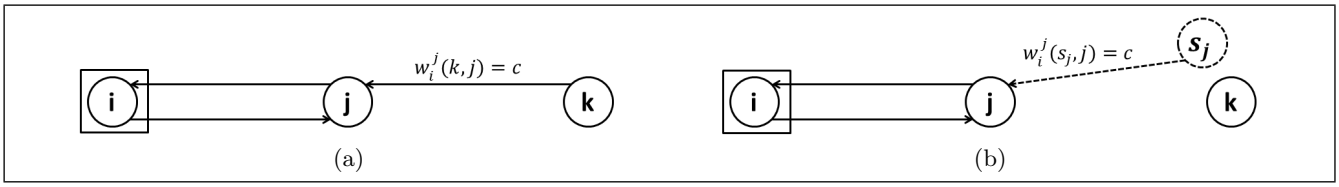

Figure 6: An illustration of the strongly beneficial sybil attack used in the proof for Theorem 1.

3.2 (Im-)Possibility of K-Sybil-Proofness

In this section we explore whether we can achieve any kind of formal sybil-proofness guarantees, despite the strong negative results from the last section. The only property that we can reasonably relax for the design of useful accounting mechanisms is the single-report responsiveness property. We can conceive of mechanisms that require two, or more generally, K, positive or negative reports about an agent, before the mechanism assigns a score distinct from S_0^M to that agent. This leads to a generalization of the responsiveness property and a corresponding notion of sybil-proofness.

Definition 15. **(K-Report Responsiveness Property)** An accounting mechanism M has the K-report responsiveness property if, for any agent i, there exists a subjective work graph $G_i = (V_i, E_i, w_i)$ and choice set C_i, with node l and a set of nodes $V_K \subseteq V_i$ with $|V_K| = K$, such that nodes i and all nodes in V_K are neighbors in G_i and no path is connecting i and l, and there exists a graph $G_i' = (V_i', E_i', w_i')$ with $V_i' = V_i$, $E_i' = E_i \cup \{(k,j),(j,k)|k \in V_K\}$, and $w_i'(e) = w_i(e)$ for all $e \in E_i \setminus \{(k,j),(j,k)|k \in V_K\}$, and there exists a constant $c \in \mathbb{R}_{>0}$ with $w_i'^j(k,j) = c$ for all $k \in V_K$, such that:

$$S_{ik}^M(G_i', C_i') > S_0^M.$$

Definition 16. **(K-Sybil-Proofness)** An accounting mechanism M is *K-Sybil-proof* against strongly (weakly) beneficial sybil attacks if, for every work graph G_i, there exists no passive or active strongly (weakly) beneficial sybil attack with K or fewer sybils for M.

THEOREM 2. *For every accounting mechanism that satisfies independence of disconnected agents, is symmetric, and has the K-report responsiveness property, there exists an active weakly beneficial sybil attack.*

PROOF. Assume that accounting mechanism M is K-report responsiveness. Then there exists a subjective work graph G_i and nodes l and V_K as described in Definition 15. If all agents in V_K make a report about their edge to l with weight c leading to G_i', then the resulting score for agent l is greater than S_0^M; in particular, we assume that $A(G_i', C_i') = l$. If we remove one agent k^* from the set V_K this leads to G_i'' where $S_{il}^M(G_i'', C_i'') = S_0^M$, and we assume that now $A(G_i'', C_i'') \neq l$. Now we let agent j create a sybil agent s_j. Because of the independence of disconnected agents, this does not change any of the scores. Next, we assume that s_j takes the place of k^*, performing and consuming the same amount of work as k^* had such that, from i's perspective, agents k^* and s_j look the same. Because M is symmetric, if agent s_j now makes a report about edge (l, s_j) with weight c, then $S_{il}^M(G_i''', C_i''') > S_0^M$. W.l.o.g. we can assume that $l = j$. Thus, this constitutes an active weakly beneficial sybil attack. □

We will now show how to turn any reasonable accounting mechanism into a K-report responsive mechanism that is K-sybil-proof against strongly beneficial sybil attacks:

Definition 17. **(K-Elimination-Wrapper)** A K-Elimination-Wrapper W takes as input an accounting mechanism M, a subjective work graph $G_i = (V_i, E_i, w_i)$, and a choice set C_i, and determines the scores $S_{ij}^W(M, G_i, C_i)$ for each agent $j \in C_i$, as viewed by agent i. Let $\mathcal{P}(V_i)$ denote the powerset of V_i, and let $\mathcal{P}_{\leq K}(V_i)$ denote the set of subsets of $\mathcal{P}(V_i)$ of cardinality less than or equal to K. We let $G_i \setminus X$ denote the subjective work graph that results from taking G_i and removing all nodes in X from V_i. Then:

$$S_{ij}^W(M, G_i, C_i) = \min_{X \in \mathcal{P}_{\leq K}(V_i)} \{S_{ij}^M(G_i \setminus X, C_i \setminus X)\}.$$

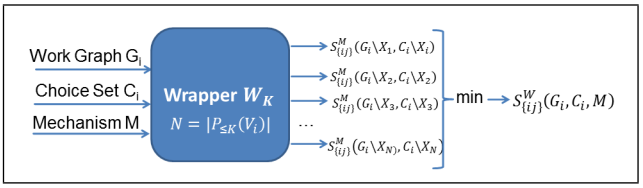

Figure 7: The K-Elimination-Wrapper.

So far, we have not specified how an accounting mechanism uses a subjective work graph to compute the scores, in particular, how indirect reports from other agents are used. The design of accounting mechanisms is based on the notion of transitive trust, i.e., if A trusts B and B trusts C, then A also trusts C to some degree. Furthermore, we assume that the only way to *earn trust* is by performing work. To obtain a mechanism that is sybil-proof against certain attacks, we assume that an agent's maximum influence on another agent's score is bounded. We let $work(j)$ denote the total amount of work that agent j has performed in the network. Here, j represents any attacking agent which can be a real agent or a sybil node:

Definition 18. **(Work-monotonic Transitive Trust Property)** An accounting mechanism M has the work-monotonic transitive trust property, if for any subjective work graph G_i and choice set C_i with $j \in V_i$ and $k \neq j \in C_i$, and any misreport attack σ_j such that $G_i' = G_i \downarrow \sigma_j$:

$$|S_{ik}^M(G_i, C_i) - S_{ik}^M(G_i', C_i)| \leq f(work(j)) \in \mathbb{R}_+ \setminus \{\infty\}$$

and $f(0) = 0$ and $f(\cdot)$ is weakly monotonically increasing.

In words, this assumes that the maximum difference in agent k's scores that j can cause via a misreport is bounded by a finite function that weakly monotonically increases in the amount of work performed by j and that is zero when $work(j) = 0$.

THEOREM 3. *A K-elimination-wrapper applied to an accounting mechanism that satisfies the work-monotonic transitive trust property leads to an accounting mechanism that is K-sybil-proof against strongly beneficial passive or active sybil attacks.*

PROOF. Assume there exists a sybil attack by some agent j that involve less than or equal to K sybils and is strongly beneficial. The K-elimination wrapper iteratively removes all subsets of agents of size K or less from G_i, computes all scores without those subsets, and ultimately takes the minimum. Thus, in one of those iterations, all of j's sybils will be removed from G_i, and the resulting score will be part of the overall minimization of the wrapper. Thus, if an agent's score before the sybil attack was lower than afterwards, then the wrapper will take the score from before the attack. This excludes options (1) and (3) from the set of beneficial sybil attacks (see Definition 9) where the goal of the sybil attack was to increase agent j's or one of the sybils' scores. This only leaves options (2) and (4). Consider first option (2), where the other agents' scores are lowered enough such that afterwards j is allocated. If j is allocated, j is inside the choice set and thus the attack must have been performed by one of j's sybils, let it be s. Because the mechanism satisfies the work-monotonic transitive trust property, we know that to decrease an agent's score, s must have performed work before. Furthermore, the amount of decrease is limited by $f(work(s))$. Now, if the attack is successful, then j gets to consume some units of work "for free". However, after consuming a certain amount of work, j's score is lowered again, and at some point, j's score will be lower than k's score again. Thus, now another sybil attack would be necessary, which again would require the sybil agents to perform work. Thus, the amount of free work resulting from this sybil attack is bounded: for every x units of "free" work, the sybil attack requires a certain fixed amount of work as well. The argument is analogous for option (4). Thus, the sybil attack can be weakly beneficial, but not strongly beneficial. □

4. DISCUSSION AND FUTURE WORK

We now discuss some practical aspects related to the theoretical findings from the previous section.

K-Sybil-Proofness in Practice.
Note that using the K-elimination-wrapper does not provide robustness against weakly beneficial sybil attacks. Furthermore, even achieving K-sybil-proofness against strongly beneficial sybil attacks comes at a cost: the resulting mechanism is only K-report-responsive and ignores a larger part of the available information compared to a single-report responsive mechanism. Assuming random interactions between peers, the probability of having K reports about an agent decreases exponentially in K. Thus, real-world system designers face an important trade-off between (limited) robustness against sybil attacks on the one side, and informativeness on the other side. In some domains, creating one or two sybils may be relatively cheap, but creating more sybils could become very expensive. For example, you might have one home IP address and one work IP address that you control, and obtaining a third IP address might be reasonably costly. Thus, for this particular domain, a 2-sybil-proof mechanism might provide useful robustness.

In some domains the interactions between peers are not random, e.g., in P2P file sharing communities where agents may have similar taste preferences. In these highly clustered domains, it might be reasonable to assume that each agent has an average of more than K reports about each other agent, such that after applying a K-elimination-wrapper, the resulting scores would still be informative enough. In future work, we will analyze this trade-off in more detail.

Max-Flow: Robustness to Sybil Attacks.
We have seen that fully sybil-proof accounting mechanisms do not exist, and even limited robustness comes at a high price. One way to address this problem in practice is the application of the max-flow algorithm inside an accounting mechanism. In Seuken et al. [11], we present mechanisms that compute agent j's score from i's perspective as $MF(j,i) - MF(i,j)$, where $MF(x,y)$ denotes the maximum flow on agent i's subjective work graph. While this does not provide any formal guarantees against sybil attacks, using max-flow provides additional robustness in practice: it bounds the influence of any agent by the total amount of work performed by that agent itself, which is a special form of the work-monotonic transitive trust property. This limits the power of sybil attacks, making them more costly and thus less attractive for the attacking agent. By the same argument, max-flow is also useful to protect against Byzantine agents, i.e., agents that try to harm the network or specific agents in the network. For example, if a Byzantine agent reports that agent i has consumed 1,000,000 units of work from him, and if other agents believe this report, then agent i will be unable to receive any work from those agents in the future. Using max-flow makes Byzantine attacks more difficult and costly for the attacking agent, thereby providing significant robustness against such attacks in practice.

5. CONCLUSION

In this paper, we have studied the (non-)existence of sybil-proof accounting mechanisms. Our main result is that under reasonable assumptions, no sybil-proof accounting mechanism exists. This is a strong impossibility result, answering an important open question. This result is also in stark contrast to well-known positive results regarding sybil-proofness reputation mechanisms. We have explored a significantly weaker notion of sybil-proofness, where we have shown that by using a K-elimination wrapper, we can design accounting mechanisms that are K-sybil-proof against strongly beneficial sybil manipulations. However, these mechanisms are still susceptible to weakly beneficial attacks. It is noteworthy that all of our results hold independent of the behavior of other agents in the network. Finally, we have illustrated the benefits of using max-flow algorithms for the internal score computation. While this does not provide formal sybil-proofness guarantees, it limits the influence of any particular agent, thereby providing some robustness against sybil manipulations and Byzantine attacks in practice. We hope that this work will inform others about the general limitations of the design of accounting mechanisms for distributed work systems and encourage new pragmatic analyses.

Acknowledgements

We thank Michel Meulpolder, Johan Pouwelse, Ian Kash, Ariel Procaccia, Mike Ruberry, and three anonymous reviewers for helpful feedback on this work. The first author was supported by a Microsoft Research PhD Fellowship. This work is supported in part by NSF grant CCF-0915016.

6. REFERENCES

[1] A. Cheng and E. Friedman. Sybilproof Reputation Mechanisms. In *Proceedings of the ACM SIGCOMM Workshop on Economics of Peer-to-Peer Systems (P2PECON)*, pages 128–132, Philadelphia, PA, August 2005.

[2] A. Cheng and E. Friedman. Manipulability of PageRank under Sybil Strategies. In *Proceedings of the 1st Workshop of Networked Systems (NetEcon06)*, Ann Arbor, MI, June 2006.

[3] B. Cohen. Incentives Build Robustness in BitTorrent. In *Proceedings of the Workshop on Economics of Peer-to-Peer Systems (P2PEcon)*, Berkeley, CA, June 2003.

[4] M. Feldman, K. Lai, I. Stoica, and J. Chuang. Robust Incentive Techniques for Peer-to-Peer Networks. In *Proceedings of the 5th ACM Conference on Electronic Commerce (EC)*, New York, NY, May 2004.

[5] M. Feldman, C. Papadimitriou, J. Chuang, and I. Stoica. Free-Riding and Whitewashing in Peer-to-Peer Systems. *IEEE Journal on Selected Areas in Communications*, 24(5):1010–1019, 2006.

[6] E. Friedman, P. Resnick, and R. Sami. Manipulation-Resistant Reputation Systems. In N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, editors, *Algorithmic Game Theory*, pages 677–698. Cambridge University Press, New York, NY, 2007.

[7] M. Meulpolder, J. Pouwelse, D. H. Epema, and H. J. Sips. BarterCast: A Practical Approach to Prevent Lazy Freeriding in P2P Networks. In *Proceedings of the 6th International Workshop on Hot Topics in Peer-to-Peer Systems (Hot-P2P)*, Rome, Italy, May 2009.

[8] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford Digital Library Technologies Project, 1998.

[9] M. Piatek, T. Isdal, A. Krishnamurthy, and T. Anderson. One Hop Reputations for Peer to Peer File Sharing Workloads. In *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, pages 1–14, San Francisco, California, April 2008.

[10] P. Resnick and R. Sami. Sybilproof Transitive Trust Protocols. In *Proceedings of the 10th ACM Conference on Electronic Commerce (EC)*, Stanford, CA, July 2009.

[11] S. Seuken, J. Tang, and D. C. Parkes. Accounting Mechanisms for Distributed Work Systems. In *Proceedings of the Conference on Artificial Intelligence (AAAI)*, Atlanta, GA, July 2010.

[12] D. Sheldon and J. Hopcroft. Manipulation-Resistant Reputations Using Hitting Time. In *Proceedings of the 5th Workshop on Algorithms for the Web-Graph (WAW)*, San Diesgo, December 2007.