The Sybil Attack - Theory and Practice

Kelong Cong
Delft University of Technology
k.cong@student.tudelft.nl

ABSTRACT

TODO

1. INTRODUCTION

Electronic commerce and online social networks are common phenomenons at the present time. They allow us to orchestrate many aspects of our lives in the comfort of our homes, behind the monitors of our devices. An online identity is often required to use such services, for examples we must create an account to tweet[1] our friend, who must also have an account. In this scenario, users can choose to remain psudonymous if they are careful, where their real-life identity is uncorrelated with their online identity.

While creating pseudonyms is useful for protecting users' privacy, it also opens an alleyway for attackers. The Sybil attack, first described by Douceur[23], is an attack where an entity can assume multiple identities or Sybils, and then attack either another entity or undermine the whole system. For example, a malicious Twitter user can create many fake identities and have the fake identities follow his real identity, thus creating a false reputation. It is one of the most important attacks because it leads to a large number of consequences including but not limited to spreading false information, identity theift[8] and ballot stuffing[7]. Furthermore, to the best of our knowledge, there is no general solution for preventing the Sybil attack.

In this work, we survey various aspects of the Sybil attack. But in contrast with previous surveys, we include both the theoretical and practical aspects. First, we describe the Sybil attack in more detail and and illustrate its importance by looking at how researchers and black-hat hackers mounted the attack on real-world e-commerce and online social network systems in section 3. Since there is a large variety of Sybil attack defence mechanisms, from using trusted-third-party to exploiting the graph characteristics in online social networks, thus we classify these mechanisms by their "main idea" in section 5. Finally we present the related work and conclude in section 6 and section 7.

[1]A message sent using Twitter is a tweet.

© 2016 ACM. ISBN 123-4567-24-567/08/06.
DOI: 10.475/123_4

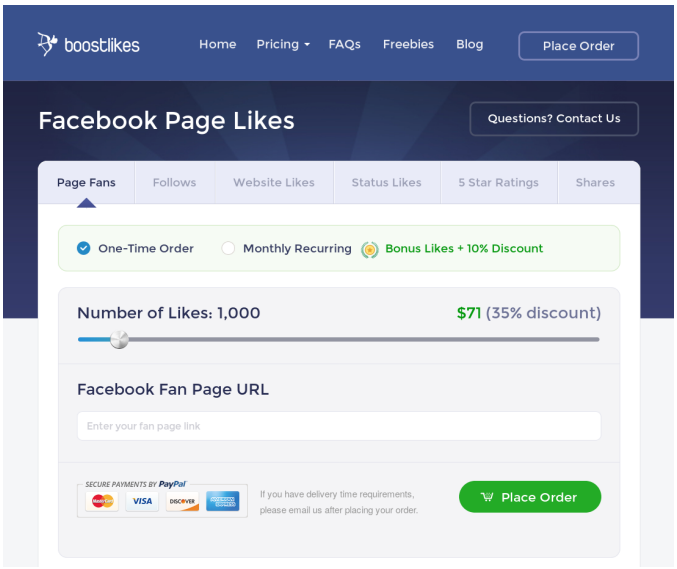

Figure 1: Screenshot of the Facebook likes service page of boostlikes.com.

2. MOTIVATION

We begin our survey by showing some alarming Sybil attacks happening in the real-world. Social network and micro-blogging websites are popular platforms for organisations to improve public relations and their reputation, but they are also platforms to spread propoganda. A recent article in the Atlantic described how Twitter bots (Sybils) are shaping the 2016 US presidential election[32]. Over a third of pro-Trump tweets and almost a fifth of pro-Clinton tweets, totalling at about 1 million, came from bots. The article questions whether the bots are a threat to democracy because opinions of real users are eclipsed by spam of bots.

Using Sybils to manipulate public opinion is not only accessible to campaigners with a large budget. There are marketplaces where anybody can purchase reputation scores such as Twitter followers. BoostLikes shown on Figure 1 is a professionally presented website, it offers a large range of services including Facebook likes, Twitter followers, Instagram followers and YouTube views. SocialFormulae (Figure 2) is a similar service but at a much lower price point, one thou-

Figure 2: Screenshot of the main banner on socialformulae.com.

sand Twitter followers is only $9.99. There can be little doubt that those companies use automated bots to provide their services.

SadBotTrue and its related website Socialpuncher publishes studies on social media fraud. Two of their studies is particularly useful for demonstrating the scale of the Sybil attack on Twitter. Firstly, there exist a botnet that consist of 3 million accounts. Since their creation, they generated 2.6 billion tweets. Surprisingly, all of the 3 million accounts were created on the same day and the account names are simply numbered sequentially[71]. Such an obvious activity should be easily detectable by Twitter, but these accounts are still not closed at the time of writing. Secondly, the top-100 Twitter users have 523 million unique followers between them, but 310 million are bots, that is almost 60%[80]. Suppose the bots all belong to the same attacker, then they can effectively suppress the opinions of the real users.

Clearly, the defence mechanisms employed by social network and micro-blogging websites are not adaquate to combat the Sybil attack. If the Sybils infiltrate even more of our cyberspace, then it may become a form of censorship. In that case, can we still be considered to have the right to freedom of speech?

3. THE SYBIL ATTACK

The Sybil attack is coined by Douceur[23] in 2002 in the context of peer-to-peer systems. In this section, we introduce the key theoretical results and the definitions used in the remainder of this survey.

3.1 Theoretical Results

Douceur defined the Sybil attack as forging multiple identities under the same entity[23]. An entity can be for example a physical user of the system and identities are how entities present themselves to the system. Thus a local entity has no direct knowledge of remote entities, only their identities. The forged identities do not necessarily follow the protocol specified by the underlying network, i.e. they assume the characteristics of Byzantine fault[48].

The author modelled the system as a general distributed computing environment where there is no constraint on the topology, every node has limited computational resources and messages are guaranteed to be delivered. Under this model, the author proved that the Sybil attack is always possible without a central, trusted authority.

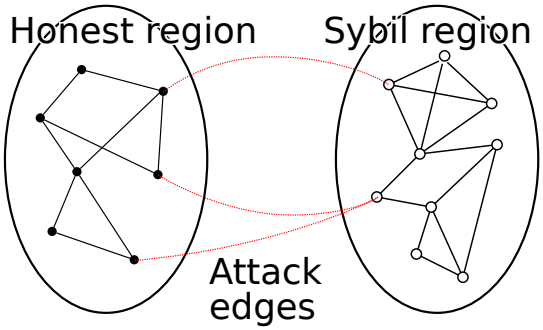

Figure 3: Visualisation of the Sybil attack in online social networks

Cheng and Friedman proved a similar result in the context of reputation systems[12]. Reputation systems are commonly used in e-commerce websites and the internet in general, where identities are rewarded by their good behaviour or usefulness. Google's PageRank[64] is an example of a reputation system, where a large number of links to a website makes it more reputable. It was formally proven that peer-to-peer reputation systems cannot be made to prevent the Sybil attack, it is only possible prevent it by using trusted parties.

3.2 Model and Definitions

One of the common models, especially in the context of online social networks, is shown in Figure 3. It is first introduced by the authors of SybilGuard[103]. Nodes inside the left region are identities created by benevolent entities, the edges connecting those nodes are real-world trust relationships. The right region contains the Sybils and they are connected with fake relationships. The edges connecting the two regions are called *attack edges*. These can be created by tricking a benevolent user to befriend a Sybil, stealing a benevolent user's account and so on. Many Sybil defence mechanisms rely on the fact that attack edges are difficult to create as we will describe in section 5.

4. SYBIL ATTACK IN VARIOUS APPLICATIONS

Sybil attacks can be mounted in different applications and

cause a large array of consequences. This section categorises the attacks by the goal for four common applications. (1) P2P (peer-to-peer) file sharing networks such as BitTorrent, (2) OSN (online social networks) such as Twitter and Facebook, (3) reputation systems such as eBay and (4) WANET (wireless ad-hoc networks) such as sensor networks. We hope this section further illuminates the alarming consequences of the Sybil attack.

4.1 The Sybil Attack in P2P FIle Sharing Networks

P2P (peer-to-peer) file sharing networks are distributed computer networks that are built for discovering and sharing files. BitTorrent[14] is likely the most popular P2P network at the time of writing. Due to their open and distributed nature, they are vulnerable to the Sybil attack.

4.1.1 Denial of Service

By exploiting vulnerabilities in the BitTorrent network, denial of service attack can be directed at any machine connected to the internet, not just machines in the network[78]. The main idea is to report the victim as the tracker (a server that coordinates the peers). El Defrawy, Gjoka and Markopoulou created a small scale proof-of-concept attack. Using only one machine, they could generate enough traffic to cripple small organisations and home users. The authors suggested that if Sybils are created to perform the same attack aimed at a single victim, then it could easily throttle links with much higher bandwidth[24].

4.1.2 Index Poisoning

P2P networks often implement a DHT (distributed hash table). The DHT in BitTorrent is called Mainline-DHT, based on Kademlia[58]. Keys are the infohashes (file identifiers) and values are the metadata of the files, these are distributed across all the participating peers. Every node stores a routing table and requests are routed iteratively to the node responsible for a particular key[53]. The goal of index poisoning is to corrupt routing table so that honest peers fail to find the values they want. It can be mounted by injecting Sybils into the DHT that do not follow the protocol. Wang and Kangasharju created honeypots in the BitTorrent network and detected as many as 300,000 Sybils[95]. Similar attacks are possible in other P2P networks such as Overnet[52].

4.1.3 Eclipse Attack

4.2 The Sybil Attack in OSN

OSN (online social networks) are vulnerable to the Sybil attack even when most of them use a central, trusted authority such as Facebook. In OSN, users create profiles and form relationships with friends. In contrast with real world relationships, it is much easier to create relationships in OSN even with strangers. In 2008, Sophos conducted an experiment where they created a Facebook profile and send friend requests to 200 random users, and 41% of the users accepted the friend request[82]. A report by Facebook at the end of 2011 stated 5-6% of their accounts are fake[65]. Combining with the ability to create new identities with very little cost, it is possible to perform many types of attacks which we outline below.

Note that online social networks often have a reputation aspect as well, for example a Facebook page with a lot of fans may be considered to be more reputable than others. We discuss attacks specific to OSN in this section and attacks on reputation in subsection 4.3

4.2.1 Identity Theft

Authors of [8] created two attacks - profile cloning and cross-site profile cloning, targeting five social network sites including Facebook and LinkedIn. The iCloner system was created to automate these attacks.

In profile cloning, iCloner uses publicly available information to automatically create clones of the victim's profiles, effectively creating Sybils. iCloner then sent friend requests from the cloned profile to the friends of the victim. The fact that the victim may have many friends that they do not contact very often, e.g. friend from primary school living in another country, makes this attack highly effective. The authors found that the acceptance rate for cloned profiles was over 60%. Much higher than the acceptance rate of 30% for fictitious profiles. Once the friendship is established, it is possible to extract private information that is not available publicly and perform identity theft.

The idea of cross-site profile cloning is similar, except the cloned profile is created on another social network site that the victim does not yet use. Once the cloned profile is created, iCloner attempts to identify friends of the victim and begins sending friend requests. Similarly, 56% of the friend requests were accepted.

A more recent study created SbN (Socialbot Network) targeting Facebook[9]. Each socialbot is a Sybil created by the attack, it controls a forged profile and minic human behaviour to avoid detection. The attacker is the botmaster who coordinates the socialbots to achieve a common objective such as infiltrating the target OSN by creating friend relationships with real users. The authors found that infiltration success rate was as high as 80% and the FIS[84] (Facebook Immune System) was not sufficient to prevent the attack. Once the relationships are established, the botmaster can command the socialbots to start gathering private information which can then be used for identity theft.

These examples demonstrate that the carelessness of users and the ability to create Sybils makes OSN vulnerable to identity theft. Moreover, identity theft is only an entry point. Once trust relationships are established, the attacker can perform many other types of attacks such as spamming, phishing or astroturfing to gain advantage.

4.2.2 Astroturf

Astroturfing is an act of creating grassroot movement that are in reality carried out by a single entity, effectively spreading misinformation to legitimate users. It relies on the ability to create Sybils in the underlying social network. This type of attack is especially effective in social networks such as Twitter where a lot of the social interaction such as sending messages happen in the public.

In the 2010 Massachusetts senate race, Mustafaraj and Metaxas found evidence that Republican campaingners created fake Twitter accounts and used them to send spam. The spam caused Google real-time search results to tip in their favour thus causing a spread of misinformation[61]. Ratkiewicz et el. suggest that this type of attack can be mounted cheaply and may have a larger influence than traditional adversiting[67].

The Truthy system[67] is a web service that perform real-time analysis of Twitter to detect political astroturfing. In the 2010 U.S. midterm election, the authors found accounts which generated a lot of retweets but no original tweets. More importantly, they uncovered a network of bot accounts that injected thousands of tweets to smear the Democratic candidate.

In 2012, Wang et el. investigated two of the largeset crowdturfing[2] platforms in China that brings together buyers and sellers - Zhubajie and Sandaha. One of their services is perform astroturfing on Weibo (The Chinese Twitter). The authors found that the 5364 sellers collectively own 14151 Weibo accounts and the top 1% of the sellers own over 100 accounts. Furthermore, the business is growing and more than $4 million have been spent on these two platforms over five years[94].

4.2.3 Spam

Spamming, much like in the context of email, is the act of sending unsolicited or undesire messages (spam). The goal of the attacker varies from advertisement to phishing and spreading malicious software[36, 86]. Many studies have characterised the behaviour of the spammers and found that they either use fake accounts or stolen accounts[85, 100, 30]. Some authors have worked with the service provide to close the spam accounts, but it is clearly not sufficient as we described in section 2.

4.3 The Sybil Attack in Reputation Systems

Reputation systems cultivate collaborative behaviour by allowing entities to trust each other based on community feedback, usually in the form of a reputation score. Entities decide whom to trust based on the reputation scores, thus entities are also incentivised to behave honestly. Reputation systems are found in many context. In e-commerce, namely eBay, researchers found that the merchant's reputation "is a statistically and economically significant determinant of auction prices"[39], and "buyers are willing to pay 8.1% more" for goods sold buy a reputable merchant[68]. The file sharing peer-to-peer network BitTorrent uses tit-for-tat as an ephemeral reputation system to encourage peers to upload in exchange for better download speeds[13]. The aforementioned PageRank[64] is also a reputation system, used for ranking reputable websites higher in Google's search results.

Unfortunately, reputation systems are also vulnerable to the Sybil attack. Worryingly, there appears to an industry built around it, and their products are easily accessible in the clearnet. In this section, we describe practical attacks on reputation systems.

4.3.1 Self-promoting

In self-promotion, the goal of the attacker is to illegitimately raise its own reputation. A common way to perform self-promition is to create Sybils and have them create positive reputation for the attacker's main identity.

Dini and Spagnolo studied the economics of buying reputation on eBay. The authors discovered many cheap items (around €0.7) for sell are simply there to boost feedback. For example, one of the item is titled "Apple Cranberry Crisp Recipe + 100% Positive Feedback". The authors successfully boosted their feedback by purchasing such items. But they made an unsuccessful attempted to place a bid on their own good with a fake account[22].

De Cristofaro et el. performed an emperical study on Facebook page promotion using like farms[18]. Some of the farms such as `SocialFormulae.com` are clearly operated by bots and the operator does not attempt to hide it, others such as `BoostLikes.com` tries to minic human users. The authors purchased the "1000 likes" service on their empty Facebook pages. In under a month, many empty pages have accumulated almost 1000 likes as promised by the like farms. The authors empty accounts were not terminated. Only a small number of the liker's account were terminated.

SEOClerks and MyCheapJobs are also evidences of marketplaces for self-promotion. Some of the top services include "1 million Twitter followers" at $849, "1000+ Instagram followers" at $10 and so on. The revenues of those two marketplaces are estimated to be at $1.3 million and $116 thousand, respectively[25]. Although the authors did not investigate the properties of the fake followers, there is little doubt that many of accounts used in these services are Sybils.

4.3.2 Slandering

The goal of a slandering attack is to illegitimately produce negative feedback to undermine the reputation of the target. It is easy to imagine the improvement in effectiveness when using multiple Sybils. From the best of our knowledge, there are no published studies on real-world slandering. But research has shown having a negative feedback may harm the target's ability to do business[5].

4.3.3 Whitewashing

In whitewashing, attackers abuse the reputation system for temporary gain and then escape the consequences by joining the reputation system under a new identity to shed their bad reputation. Clearly, whitewashing is only possible when the Sybil attack is possible. Again, there are no studies on whitewashing in the real-world. But many have suggested that it is feasible attack[38, 57].

4.4 The Sybil Attack in WANET

WANET (wireless ad-hoc networks) is a dynamic, self-configuring, self-healing wireless network. Ad-hoc in this case means it does not rely on existing infrastructure for the network to function. Each node in the network is responsible for some general tasks such as routing, and some application specific tasks such as gathering data from its sensors in the case of a sensor network.

Akin to the other applications, an attacker in a WANET may own a single physical node, but it may behave as if it were a large number of nodes. Many WANET designs involve a reputation system[29, 10], thus the same attacks from subsection 4.3 applies here. In this section we describe the WANET specific attacks. From the best of our knowledge WANET are not widely deployed in practice, thus there is little research on real-world attacks.

4.4.1 Unfair Resource Allocation

Nodes in WANET often have limited resources such as bandwidth of the radio channels. Resources such as these must be shared between the neighbours using time slices. When the neighbours are Sybils, then the attacker can receive an unfair amount of resource allocation and denies re-

[2]Crowdsourced astroturfing.

sources for the honest nodes[63]. In contrast with the other attacks, this works even when the Sybils are not behaving malicioiusly.

4.4.2 Routing Disruption

An important routing technique is multipath routing, data is routed using multiple paths in the network for better fault-tolerance and bandwidth. However, if Sybils are present in the network, then the different paths may in fact go through the Sybils owned by the same attacker. Another technique geographic routing, nodes route data depending on the geographic location of their neighbours. Sybils in the network can be in more than one place at a time, thus significant distrupting the routing algorithm[44].

4.4.3 Spreading False Information

Nodes often need to exchange information with each other to satisfy the underlying requirements of the application. Some of the common tasks include data aggregation, voting. With enough Sybils, it is possible to manipulate the aggregated data or the poll to benefit the attacker. For example, sensor networks may use a bollot to detect misbehaving nodes, the attack could use its Sybils to claim that a honest node is misbehaving and have the other nodes expel it from the network[63].

4.5 TODO

a test bed for sybil attacks[40]
Quantifying Sybil attack[55]

5. DEFENCES

In this section we categorise various defence techniques against the sybil-attack. Many of them are independent of the application, thus we classify them on their main idea, and state explicitly when the mechanism is application specific.

5.1 Certificate Authority

CA (certificate authorities) check the users' identities and then issues certificates to benevolent users. The certificate can be tangible (trusted hardware[63]) or non-tangible (public key certificate) depending on the application. When an identity wishes to user the application, the CA must verify the validity of its certificate to ensure one-to-one correspondence. This mechanism prevents the Sybil attack as long as the CA does not make mistakes in the issuance stage.

CA can prevent the Sybil attack but it also has a lot of downsides. (1) Users have different opinions and may not agree on a single CA. (2) Users living in authoritarian regimes may not have access to the CA in use. (3) It is difficult to scale up a CA to meet increasing users demands. (4) Anonymity is difficult to obtain because the CA has complete information of the entities. (5) It is a central point of failure; i.e. if the attacker obtains the private key to create certificates then he or she can easily generate Sybils, if the CA goes offline then the application ceases to function because it can no longer verify identites.

Many existing systems today use a form of CA. X.509 is a standard for certificates and is used in a large variety of applications, for example websites (TLS), email (S/MIME), smart cards and so on. Payment systems such as PayPal verifies identities using credit card billing addresses.

5.2 Resource Testing

Every attacker can create multiple Sybils, but the attack cannot duplicate its resources the same way. In resource testing, the goal is to find identities that possess fewer than the expected amout of resources. The resource type can vary depending on the application. In wireless networks it may be using radio channels[63], in online social networks it may be IP addresses[27] and solving computationally demanding puzzles in P2P networks[4]. Resource testing may deter casual attackers but its usefulness degrades for resourceful attackers.

For example, in the Tarzan P2P network, neighbours are selected not from all known IP addresses, but from distinct IP prefixes[27]. The effectiveness of the Sybil attack is reduced if the attackers cannot easily create Sybils in a large range of IP prefixes.

Another example for resource testing is self-registration[21]. When a peer wish to join the network, it needs to compute a ID which is a hash of its own IP address and port number. While participating in the nettwork, other peers need to verify that the ID matches the peer's origin.

Bitcoin is also resource testing...

5.3 Registration Fee

Using registration fee is similar to resource testing except it only happens on registration. Entities can be charged a fee for creating identities, often facilitated by a central authority. The fee need to be set appropriately so that the cost of creating Sybils outweights the benefits. The fee does not need to be monetary. For instance, CAPTCHA[92] is a form of registration fee. It prevents programs from automatically creating new identities and limits the rate at which identities can be created.

Feldman et el. proposed another form of registration fee for P2P networks - the adaptive stranger policy[26]. When new peers join the network, they are treated using a policy that is adapted from previous newcomers. For example, the new peers may be expected to contribute to the network before they are allowed to receive benefits from the "mature" peers. The downside is that the policy may deter benevolent users from joining the network in the first place.

5.4 Network Flow Based Techniques

Network flow based techniques began with BarterCast[59]. It was initially designed to combat freeriding in P2P file sharing networks, where users are selfish and do not share content, but its idea can be extended combat the Sybil attack.

The main idea comes from real-world social networks, where the reputation of a person can be from direct experiences, or information obtained frome someone else. The direct experience is always true, but the indirect information may not be, i.e. people can lie about their experiences. Humans solve the problem by treating the indirect information with a grain of salt unless the source of the information is highly trusted.

BarterCast applies this idea in P2P file sharing networks. Peers all maintain a subjective graph which is created by exchanging messages with their neighbours. The direct experiences measured by the number of bytes uploaded and downloaded are represented by the outgoing and incoming edges from the peer, respectively. Indirect experiences are represented by edges that are not directly connected to the

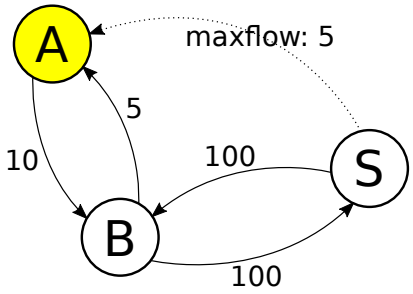

Figure 4: Subjective graph of A. The numbers are the amout of data transfered, they can be seen as the capacity in the context of the maximum flow problem.

peer. For example in Figure 4, A is the subject, it has direct experiences with B and B has told A about S, so it has indirect information about S. But A is unsure about the truthfulness of S's contribution, so it only trusts S as much as it trusts B. This idea is realised using a maximum flow algorithm and the final reputation metric is given in Equation 1.

$$R_i(j) = \frac{\arctan(\text{maxflow}(j,i) - \text{maxflow}(i,j))}{\pi/2} \qquad (1)$$

BarterCast does not prevent the Sybil attack by itself. Because attackers can first upload a lot of data to obtain a good reputation in the network. If the attacker now creates Sybils and false report of the Sybils saying that they uploaded a lot. Then the peers who have interacted with the attacker will be tricked to think that the Sybils also have a high reputation. To fix this problem, Delaviz et el. created SybilRes[19]. The main idea is the following. Suppose there are two peers A and B who are sharing data. If A is uploading (represented by an outgoing edge) to B, then it decreases the weight of the incoming edge from B. Vice versa, the weight is increased for the outgoing edge when A is downloading. The rate of change depends on the capacities of the edges and the amout of data transferred after computing the reputation. Using the definition in Figure 3, the attacker cannot built up reputation for its Sybils by uploading to peers in the benevolent region beforehand, it is now forced to keep on uploading to keep its Sybil's reputation which is a much more desirable behaviour.

Seuken et al. provided a formal model of BarterCast. They found that BarterCast is vulnerable to misreporting and proposed a solution called the DropEdge mechanism[75, 76]. DropEdge, like the name implies, drops some edges in the subjective graph that satisfies the following constraints. Suppose peer A wishes to download from peers in set C (the choice set). Then any reports received by A from $p \in C$ is dropped. Also, edges with both end points in C are also dropped from A's subjective graph. Intuitively, peers in C cannot misreport their contribution. The authors formally prove this property in their work. It may seem that the DropEdge mechanism is discarding useful information, but the authors also prove that it is robust against weakly benefician Sybils, that is Sybils that do not perform actual work for benevolent peers.

Conversely, maximum flow is dual to minimum cut, so the problem of finding Sybil can also be formulated as finding sparse cuts[3]. Kurve and Kesidis devised an algorithm for finding sparse cuts to detect Sybils[47]. Unlike the aforementioned techniques, it relies on the presence of trusted nodes.

5.5 Random Walk Based Techniques

5.6 Graph Techniques

OSN (online social networks) such as Facebook can be viewed as a graph, where the nodes on the graph are identities created by users and the edges represent trust relationships.

Gal-Oz et al. [28] communities are collection of knots, sybils can form a knot? Regret[69, 70] - information from multiple dimensions Guha 04[31] - no mention of sybil attacks or attacks in general

5.6.1 Topology

SybilGuard[103] SybilLimit[102] SybilInfer[17] SybilShield[77] - assuming sybils have bad connectedness SumUp[89] GateKeeper[90] - based on SumUp Social-network[91] - community detection

Other systems are built on top: ReDS[1] suggests to use sybilimit or sybilinfer SybilProof-DHT[50]

5.7 Reputation Transfer

Trust-transfer[73]

5.8 Cryptography Based Techniques?

Secure-Overlay[54] - ID crypto and SSS Privacy-preserving[72] - blockchain? Proof-of-stake[20] SybilConf[87]

5.9 Content Driven

[11]

5.10 Other

Parental control[88] - uses parents to "observe" find suspects, only for detection, requires a sybil-proof reputation scheme DSybil[104] - recommendation system, need historical data Symon[43] - pair peers together, likelihood for both to be sybils is low, the pair monitor each other to prevent attacks XRep 02[16] IP check, and checks digest, uses existing P2P systems like Gnutella

5.11 Distrust Relationships

VoteTrust[98]

5.12 Unsorted?

Beth and PGP limits Sybil attack to some extent by using social graphs Beth 94[6] PGP (Zimmermann) 95[108]

Yu 00[101] Lee 03[49] - uses flooding, might not be scalable, only talks about DoS Marti 04[56] ARA 05[35] - no mention of sybil, prevents freeriding, prevents short-term abuse because reputation increases gradually FuzzyTrust Song 05[81] - uses fuzzy logic P2PRep/Fuzzy 06[3] - also fuzzy, does not prevent generation of false rumors Xiong 05[97] - no mention of sybil, but tries to mitigate false information

[3]Without getting into the formal definition, the sparse cut problem is to find a partition such that the ratio between the number in the cut and the number of vertices in the smaller partition is minimised. This problem is related to minimum cut.

PowerTrust 06[107] - uses "power nodes" (from power-law), no mention of sybil, some defence against colluders

Histos and Sopras[105], doesn't really have structure? Beta[41] Gupta et al.[34]

PeerTrust[96] - DHT, used P-GRID source code, has credibility rating

PerContRep[99]

5.13 Does not handle Sybil-attack?

TrustMe[79] is a reputation that focuses on anonymity, no mention of sybil attack

H-Trust[106] does not mention sybil

Coner et al.[15] assumes clients cannot perform sybil attack

TrustGuard 05[83] - assumes it is built on secure overlay networks (sybil-proof networks)

Scrivener 05[62] - assumes ID cannot be created and discarded

6. RELATED WORK

Reputation Surveys: [57] [42] ? [38] [46] [74] ? [37]

Sybil Surveys: [51] [60] [66] [33] [45] Sok[2] but also some contribution

Other: [93]

7. SUMMARY

8. REFERENCES

[1] R. Akavipat, M. N. Al-Ameen, A. Kapadia, Z. Rahman, R. Schlegel, and M. Wright. ReDS: A framework for reputation-enhanced DHTs. *IEEE Transactions on Parallel and Distributed Systems*, 25(2):321–331, 2014.

[2] L. Alvisi, A. Clement, A. Epasto, S. Lattanzi, and A. Panconesi. Sok: The evolution of sybil defense via social networks. In *Security and Privacy (SP), 2013 IEEE Symposium on*, pages 382–396. IEEE, 2013.

[3] R. Aringhieri, E. Damiani, D. Vimercati, S. De Capitani, S. Paraboschi, and P. Samarati. Fuzzy techniques for trust and reputation management in anonymous peer-to-peer systems. *Journal of the American Society for Information Science and Technology*, 57(4):528–537, 2006.

[4] J. Aspnes, C. Jackson, and A. Krishnamurthy. Exposing computationally-challenged byzantine impostors. *Department of Computer Science, Yale University, New Haven, CT, Tech. Rep*, 2005.

[5] S. Ba and P. A. Pavlou. Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS quarterly*, pages 243–268, 2002.

[6] T. Beth, M. Borcherding, and B. Klein. Valuation of trust in open networks. In *European Symposium on Research in Computer Security*, pages 1–18. Springer, 1994.

[7] R. Bhattacharjee and A. Goel. Avoiding ballot stuffing in ebay-like reputation systems. In *Proceedings of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems*, pages 133–137. ACM, 2005.

[8] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. All your contacts are belong to us: automated identity theft attacks on social networks. In *Proceedings of the 18th international conference on World wide web*, pages 551–560. ACM, 2009.

[9] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu. The socialbot network: when bots socialize for fame and money. In *Proceedings of the 27th Annual Computer Security Applications Conference*, pages 93–102. ACM, 2011.

[10] S. Buchegger and J.-Y. Le Boudec. A robust reputation system for mobile ad-hoc networks. Technical report, 2003.

[11] K. Chatterjee, L. de Alfaro, and I. Pye. Robust content-driven reputation. In *Proceedings of the 1st ACM workshop on Workshop on AISec*, pages 33–42. ACM, 2008.

[12] A. Cheng and E. Friedman. Sybilproof reputation mechanisms. In *Proceedings of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems*, pages 128–132. ACM, 2005.

[13] B. Cohen. Incentives build robustness in BitTorrent. In *Workshop on Economics of Peer-to-Peer systems*, volume 6, pages 68–72, 2003.

[14] B. Cohen. Bep 3: The bittorrent protocol specification. http://www.bittorrent.org/beps/bep_0003.html, Jan 2008. Accessed: 2016-10-20.

[15] W. Conner, A. Iyengar, T. Mikalsen, I. Rouvellou, and K. Nahrstedt. A trust management framework for service-oriented environments. In *Proceedings of the 18th international conference on World wide web*, pages 891–900. ACM, 2009.

[16] E. Damiani, D. C. di Vimercati, S. Paraboschi, P. Samarati, and F. Violante. A reputation-based approach for choosing reliable resources in peer-to-peer networks. In *Proceedings of the 9th ACM conference on Computer and communications security*, pages 207–216. ACM, 2002.

[17] G. Danezis and P. Mittal. SybilInfer: Detecting Sybil Nodes using Social Networks. In *NDSS*. San Diego, CA, 2009.

[18] E. De Cristofaro, A. Friedman, G. Jourjon, M. A. Kaafar, and M. Z. Shafiq. Paying for likes?: Understanding facebook like fraud using honeypots. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 129–136. ACM, 2014.

[19] R. Delaviz, N. Andrade, J. A. Pouwelse, and D. H. Epema. SybilRes: A sybil-resilient flow-based decentralized reputation mechanism. In *Distributed Computing Systems (ICDCS), 2012 IEEE 32nd International Conference on*, pages 203–213. IEEE, 2012.

[20] R. Dennis and G. Owenson. Rep on the Roll: A Peer to Peer Reputation System Based on a Rolling Blockchain. 2016.

[21] J. Dinger and H. Hartenstein. Defending the sybil attack in p2p networks: Taxonomy, challenges, and a proposal for self-registration. In *First International Conference on Availability, Reliability and Security (ARES'06)*, pages 8–pp. IEEE, 2006.

[22] F. Dini and G. Spagnolo. Buying reputation on eBay: Do recent changes help? *International Journal of Electronic Business*, 7(6):581–598, 2009.

[23] J. R. Douceur. The sybil attack. In *International Workshop on Peer-to-Peer Systems*, pages 251–260. Springer, 2002.

[24] K. El Defrawy, M. Gjoka, and A. Markopoulou. BotTorrent: Misusing BitTorrent to Launch DDoS Attacks. *SRUTI*, 7:1–6, 2007.

[25] S. Farooqi, M. Ikram, G. Irfan, E. De Cristofaro, A. Friedman, G. Jourjon, M. A. Kaafar, M. Z. Shafiq, and F. Zaffar. Characterizing Seller-Driven Black-Hat Marketplaces. *arXiv preprint arXiv:1505.01637*, 2015.

[26] 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*, pages 102–111. ACM, 2004.

[27] M. J. Freedman and R. Morris. Tarzan: A peer-to-peer anonymizing network layer. In *Proceedings of the 9th ACM conference on Computer and communications security*, pages 193–206. ACM, 2002.

[28] N. Gal-Oz, E. Gudes, and D. Hendler. A robust and knot-aware trust-based reputation model. In *IFIP International Conference on Trust Management*, pages 167–182. Springer, 2008.

[29] S. Ganeriwal, L. K. Balzano, and M. B. Srivastava. Reputation-based framework for high integrity sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 4(3):15, 2008.

[30] C. Grier, K. Thomas, V. Paxson, and M. Zhang. @ spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security*, pages 27–37. ACM, 2010.

[31] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th international conference on World Wide Web*, pages 403–412. ACM, 2004.

[32] D. Guilbeault and S. Woolley. How twitter bots are shaping the election. http://www.theatlantic.com/technology/archive/2016/11/election-bots/506072/, 11 2016.

[33] R. Gunturu. Survey of Sybil attacks in social networks. *arXiv preprint arXiv:1504.05522*, 2015.

[34] M. Gupta, P. Judge, and M. Ammar. A reputation system for peer-to-peer networks. In *Proceedings of the 13th international workshop on Network and operating systems support for digital audio and video*, pages 144–152. ACM, 2003.

[35] M. Ham and G. Agha. ARA: A robust audit to prevent free-riding in P2P networks. In *Fifth IEEE International Conference on Peer-to-Peer Computing (P2P'05)*, pages 125–132. IEEE, 2005.

[36] Help Net Security. Twitter accounts spreading malicious code. https://www.helpnetsecurity.com/2010/12/03/twitter-accounts-spreading-malicious-code/, 12 2010. Accessed: 2016-11-2.

[37] F. Hendrikx, K. Bubendorfer, and R. Chard. Reputation systems: A survey and taxonomy. *Journal of Parallel and Distributed Computing*, 75:184–197, 2015.

[38] K. Hoffman, D. Zage, and C. Nita-Rotaru. A survey of attack and defense techniques for reputation systems. *ACM Computing Surveys (CSUR)*, 42(1):1, 2009.

[39] D. Houser and J. Wooders. Reputation in auctions: Theory, and evidence from eBay. *Journal of Economics & Management Strategy*, 15(2):353–369, 2006.

[40] A. A. Irissappane, S. Jiang, and J. Zhang. Towards a comprehensive testbed to evaluate the robustness of reputation systems against unfair rating attack. In *UMAP Workshops*, volume 12, 2012.

[41] A. Jøsang and R. Ismail. The beta reputation system. In *Proceedings of the 15th bled electronic commerce conference*, volume 5, pages 2502–2511, 2002.

[42] A. Jøsang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. *Decision support systems*, 43(2):618–644, 2007.

[43] B. Jyothi and J. Dharanipragada. Symon: Defending large structured p2p systems against sybil attack. In *2009 IEEE Ninth International Conference on Peer-to-Peer Computing*, pages 21–30. IEEE, 2009.

[44] C. Karlof and D. Wagner. Secure routing in wireless sensor networks: Attacks and countermeasures. *Ad hoc networks*, 1(2):293–315, 2003.

[45] D. Koll, J. Li, J. Stein, and X. Fu. On the state of OSN-based Sybil defenses. In *Networking Conference, 2014 IFIP*, pages 1–9. IEEE, 2014.

[46] E. Koutrouli and A. Tsalgatidou. Taxonomy of attacks and defense mechanisms in P2P reputation systems-Lessons for reputation system designers. *Computer Science Review*, 6(2):47–70, 2012.

[47] A. Kurve and G. Kesidis. Sybil detection via distributed sparse cut monitoring. In *2011 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2011.

[48] L. Lamport, R. Shostak, and M. Pease. The byzantine generals problem. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 4(3):382–401, 1982.

[49] S. Lee, R. Sherwood, and B. Bhattacharjee. Cooperative peer groups in NICE. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 2, pages 1272–1282. IEEE, 2003.

[50] C. Lesniewski-Lass and M. F. Kaashoek. Whanau: A sybil-proof distributed hash table. NSDI, 2010.

[51] B. N. Levine, C. Shields, and N. B. Margolin. A survey of solutions to the sybil attack. *University of Massachusetts Amherst, Amherst, MA*, 7, 2006.

[52] J. Liang, N. Naoumov, and K. W. Ross. The Index Poisoning Attack in P2P File Sharing Systems. In *INFOCOM*, pages 1–12. Citeseer, 2006.

[53] A. Loewenstern and A. Norberg. Bep 5: Dht protocol. http://www.bittorrent.org/beps/bep_0005.html, Jan 2008. Accessed: 2016-10-20.

[54] E. K. Lua. Securing peer-to-peer overlay networks from sybil attack. In *Communications and Information Technologies, 2007. ISCIT'07.*

International Symposium on, pages 1213–1218. IEEE, 2007.

[55] N. B. Margolin and B. N. Levine. Quantifying resistance to the sybil attack. In *International Conference on Financial Cryptography and Data Security*, pages 1–15. Springer, 2008.

[56] S. Marti and H. Garcia-Molina. Limited reputation sharing in P2P systems. In *Proceedings of the 5th ACM conference on Electronic commerce*, pages 91–101. ACM, 2004.

[57] S. Marti and H. Garcia-Molina. Taxonomy of trust: Categorizing P2P reputation systems. *Computer Networks*, 50(4):472–484, 2006.

[58] P. Maymounkov and D. Mazieres. Kademlia: A peer-to-peer information system based on the xor metric. In *International Workshop on Peer-to-Peer Systems*, pages 53–65. Springer, 2002.

[59] M. Meulpolder, J. A. Pouwelse, D. H. Epema, and H. J. Sips. Bartercast: A practical approach to prevent lazy freeriding in p2p networks. In *Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*, pages 1–8. IEEE, 2009.

[60] A. Mohaisen and J. Kim. The Sybil attacks and defenses: a survey. *arXiv preprint arXiv:1312.6349*, 2013.

[61] E. Mustafaraj and P. T. Metaxas. From obscurity to prominence in minutes: Political speech and real-time search. 2010.

[62] A. Nandi, T.-W. J. Ngan, A. Singh, P. Druschel, and D. S. Wallach. Scrivener: Providing incentives in cooperative content distribution systems. In *Proceedings of the ACM/IFIP/USENIX 2005 International Conference on Middleware*, pages 270–291. Springer-Verlag New York, Inc., 2005.

[63] J. Newsome, E. Shi, D. Song, and A. Perrig. The sybil attack in sensor networks: analysis & defenses. In *Proceedings of the 3rd international symposium on Information processing in sensor networks*, pages 259–268. ACM, 2004.

[64] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: bringing order to the web. 1999.

[65] E. Protalinski. Facebook: 5-6% of accounts are fake. https://web.archive.org/web/20160422121639/http://www.zdnet.com/article/facebook-5-6-of-accounts-are-fake/, 2012. Accessed: 2016-10-20.

[66] G. Rakesh, S. Rangaswamy, V. Hegde, and G. Shoba. A survey of techniques to defend against sybil attacks in social networks. *International Journal of Advanced Research in Computer and Communication Engineering*, 3(5), 2014.

[67] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini, and F. Menczer. Truthy: mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th international conference companion on World wide web*, pages 249–252. ACM, 2011.

[68] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The value of reputation on eBay: A controlled experiment. *Experimental economics*, 9(2):79–101, 2006.

[69] J. Sabater and C. Sierra. REGRET: reputation in gregarious societies. In *Proceedings of the fifth international conference on Autonomous agents*, pages 194–195. ACM, 2001.

[70] J. Sabater and C. Sierra. Social regret, a reputation model based on social relations. *ACM SIGecom Exchanges*, 3(1):44–56, 2002.

[71] SadBotTrue. Chapter 32. the stealth botnet, 6 2016. Accessed: 2016-11-2.

[72] A. Schaub, R. Bazin, O. Hasan, and L. Brunie. A trustless privacy-preserving reputation system. In *IFIP International Information Security and Privacy Conference*, pages 398–411. Springer, 2016.

[73] J.-M. Seigneur, A. Gray, and C. D. Jensen. Trust transfer: Encouraging self-recommendations without sybil attack. In *International Conference on Trust Management*, pages 321–337. Springer, 2005.

[74] C. Selvaraj and S. Anand. A survey on security issues of reputation management systems for peer-to-peer networks. *Computer Science Review*, 6(4):145–160, 2012.

[75] S. Seuken and D. C. Parkes. On the Sybil-proofness of accounting mechanisms. 2011.

[76] S. Seuken and D. C. Parkes. Sybil-proof accounting mechanisms with transitive trust. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 205–212. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

[77] L. Shi, S. Yu, W. Lou, and Y. T. Hou. Sybilshield: An agent-aided social network-based sybil defense among multiple communities. In *INFOCOM, 2013 Proceedings IEEE*, pages 1034–1042. IEEE, 2013.

[78] K. C. Sia. DDoS vulnerability analysis of BitTorrent protocol. *UCLA: Techical Report*, 2006.

[79] A. Singh and L. Liu. TrustMe: anonymous management of trust relationships in decentralized P2P systems. In *Peer-to-Peer Computing, 2003.(P2P 2003). Proceedings. Third International Conference on*, pages 142–149. IEEE, 2003.

[80] Socialpuncher. How many primitive bots follow top-100? http://socialpuncher.com/top-100/how-many-primitive-bots-follow-top-100/, 9 2016. Accessed: 2016-11-2.

[81] S. Song, K. Hwang, R. Zhou, and Y.-K. Kwok. Trusted P2P transactions with fuzzy reputation aggregation. *IEEE Internet computing*, 9(6):24–34, 2005.

[82] Sophos. Sophos facebook id probe shows 41% of users happy to reveal all to potential identity thieves. https://web.archive.org/web/20140926063331/http://www.sophos.com/en-us/press-office/press-releases/2007/08/facebook.aspx, 2007. Accessed: 2016-10-30.

[83] M. Srivatsa, L. Xiong, and L. Liu. TrustGuard: countering vulnerabilities in reputation management for decentralized overlay networks. In *Proceedings of the 14th international conference on World Wide Web*, pages 422–431. ACM, 2005.

[84] T. Stein, E. Chen, and K. Mangla. Facebook immune system. In *Proceedings of the 4th Workshop on Social Network Systems*, page 8. ACM, 2011.

[85] G. Stringhini, C. Kruegel, and G. Vigna. Detecting spammers on social networks. In *Proceedings of the 26th Annual Computer Security Applications Conference*, pages 1–9. ACM, 2010.

[86] D. Tamir. Twitter malware: Spreading more than just ideas. https://securityintelligence.com/twitter-malware-spreading-more-than-just-ideas/, 4 2013. Accessed: 2016-11-2.

[87] F. Tegeler and X. Fu. SybilConf: computational puzzles for confining sybil attacks. In *INFOCOM IEEE Conference on Computer Communications Workshops, 2010*, pages 1–2. IEEE, 2010.

[88] A. Tehale, A. Sadafule, S. Shirsat, R. Jadhav, S. Umbarje, and S. Shingade. Parental Control algorithm for Sybil detection in distributed P2P networks. *International Journal of Scientific and Research Publications*, 2(5), 2012.

[89] D. N. Tran, B. Min, J. Li, and L. Subramanian. Sybil-Resilient Online Content Voting. In *NSDI*, volume 9, pages 15–28, 2009.

[90] N. Tran, J. Li, L. Subramanian, and S. S. Chow. Optimal sybil-resilient node admission control. In *INFOCOM, 2011 Proceedings IEEE*, pages 3218–3226. IEEE, 2011.

[91] B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove. An analysis of social network-based sybil defenses. *ACM SIGCOMM Computer Communication Review*, 40(4):363–374, 2010.

[92] L. Von Ahn, M. Blum, N. J. Hopper, and J. Langford. Captcha: Using hard ai problems for security. In *International Conference on the Theory and Applications of Cryptographic Techniques*, pages 294–311. Springer, 2003.

[93] D. S. Wallach. A survey of peer-to-peer security issues. In *Software Security-Theories and Systems*, pages 42–57. Springer, 2003.

[94] G. Wang, C. Wilson, X. Zhao, Y. Zhu, M. Mohanlal, H. Zheng, and B. Y. Zhao. Serf and turf: crowdturfing for fun and profit. In *Proceedings of the 21st international conference on World Wide Web*, pages 679–688. ACM, 2012.

[95] L. Wang and J. Kangasharju. Real-world sybil attacks in BitTorrent mainline DHT. In *Global Communications Conference (GLOBECOM), 2012 IEEE*, pages 826–832. IEEE, 2012.

[96] L. Xiong and L. Liu. Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities. *IEEE transactions on Knowledge and Data Engineering*, 16(7):843–857, 2004.

[97] L. Xiong, L. Liu, and M. Ahamad. Countering feedback sparsity and manipulation in reputation systems. In *Collaborative Computing: Networking, Applications and Worksharing, 2007. CollaborateCom 2007. International Conference on*, pages 203–212. IEEE, 2007.

[98] J. Xue, Z. Yang, X. Yang, X. Wang, L. Chen, and Y. Dai. Votetrust: Leveraging friend invitation graph to defend against social network sybils. In *INFOCOM, 2013 Proceedings IEEE*, pages 2400–2408. IEEE, 2013.

[99] Z. Yan, Y. Chen, and Y. Shen. PerContRep: a practical reputation system for pervasive content services. *The Journal of Supercomputing*, 70(3):1051–1074, 2014.

[100] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu. Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In *Proceedings of the 21st international conference on World Wide Web*, pages 71–80. ACM, 2012.

[101] B. Yu and M. P. Singh. A social mechanism of reputation management in electronic communities. In *International Workshop on Cooperative Information Agents*, pages 154–165. Springer, 2000.

[102] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao. Sybillimit: A near-optimal social network defense against sybil attacks. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 3–17. IEEE, 2008.

[103] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. Sybilguard: defending against sybil attacks via social networks. In *ACM SIGCOMM Computer Communication Review*, volume 36, pages 267–278. ACM, 2006.

[104] H. Yu, C. Shi, M. Kaminsky, P. B. Gibbons, and F. Xiao. Dsybil: Optimal sybil-resistance for recommendation systems. In *2009 30th IEEE Symposium on Security and Privacy*, pages 283–298. IEEE, 2009.

[105] G. Zacharia, A. Moukas, and P. Maes. Collaborative reputation mechanisms for electronic marketplaces. *Decision Support Systems*, 29(4):371–388, 2000.

[106] H. Zhao and X. Li. H-trust: A group trust management system for peer-to-peer desktop grid. *Journal of Computer Science and Technology*, 24(5):833–843, 2009.

[107] R. Zhou and K. Hwang. Powertrust: A robust and scalable reputation system for trusted peer-to-peer computing. *IEEE Transactions on parallel and distributed systems*, 18(4):460–473, 2007.

[108] P. R. Zimmermann. *The official PGP user's guide*. MIT press, 1995.