2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing

Neighbor Similarity Trust against Sybil Attack in P2P E-Commerce

Felix Musau[1], Guojun Wang[1,*], Song Guo[2], and Muhammad Bashir Abdullahi[1]
[1]School of Information Science and Engineering, Central South University
Changsha, Hunan Province, P. R. China, 410083
[2]School of Computer Science and Engineering, University of Aizu, Aizu-Wakamatsu City, Fukushima 965-8580, Japan
*Correspondence to: csgjwang@mail.csu.edu.cn

***Abstract*—Peer to peer (P2P) e-commerce applications exist at the edge of the Internet with vulnerabilities to passive and active attacks. These attacks have pushed away potential business firms and individuals whose aim is to get the best benefit in e-commerce with minimal losses. The attacks occur during interactions between the trading peers as a transaction takes place. In this paper, we propose how to address Sybil attack, which is a kind of active attack. The peers can have bogus and multiple identity to fake their own ones. Most existing work, which concentrates on social networks and trusted certification, has not been able to prevent Sybil attack peers from participating in transactions. Our work exploits the neighbor similarity trust relationship to address Sybil attack. In this approach, referred to as *SybilTrust*, duplicated Sybil attack peers can be recognized as the neighbor peers become acquainted and hence more trusted to each other. Security and performance analysis shows Sybil attack can be minimized by our proposed neighbor similarity trust.**

***Keywords*-P2P; trust; Sybil attack; collusion attack; neighbor similarity.**

I. Introduction

P2P networks range from communication systems like email and instant messaging to collaborative content rating, recommendation, and delivery systems such as YouTube, Gnutela, Facebook, Digg, and BitTorrent. They allow any user to join the system easily at the expense of trust, with very little validation control. Peers can collude and do all sorts of malicious activities. These malicious behaviors lead to service quality degradation and monetary loss among business partners. Peers are vulnerable to exploitation, due to the open and near-zero cost of creating new identities. However, if a single defective entity can present multiple identities, it can control a substantial fraction of the system, thereby undermining this redundancy [1]. The number of identities that an attacker can generate depends on the attacker's resources such as bandwidth, memory, and computational power [2]. The goal of trust systems is to ensure that honest peers are accurately identified as trustworthy and Sybil peers as untrustworthy.

In Sybil attack, malicious peers forge a lot of fake identities to disrupt the P2P e-commerce network protocols. Defending against Sybil attack is quite a challenging task. A peer can pretend to be trusted with a hidden motive. The peer can pollute the system with bogus information, which interferes with genuine business transactions and functioning of the systems [3]. This must be counter prevented to protect the honest peers. The link between a honest peer and a Sybil peer is known as an attack edge. As each edge involved resembles a human-established trust, it is difficult for the adversary to introduce an excessive number of attack edges. The only known promising defense against Sybil attacks is to use social networks to perform user admission control and limit the number of bogus identities admitted to the system [4], [5], [8], [9]. The use of social networks between two peers represents real-world trust relationship between users. In addition, authentication-based mechanisms are used to verify the identities of the peers using shared encryption keys, or location information.

Most existing work on Sybil attack makes use of social networks to eliminate Sybil attack and the findings are based on preventing Sybil identities. In this paper, we propose use of neighbor similarity trust in a grouped P2P e-commerce based on interest relationships, to eliminate maliciousness among the peers. This is refered to as *SybilTrust*. In *SybilTrust*, the interest based group infrastructure peers have a neighbor similarity trust between each other, hence able to prevent Sybil attack. *SybilTrust* gives a better relationship in e-commerce transactions as the peers create a link between peer neighbors. This provides an important manner for peers to advertise their products to other interested peers and to know new market destinations and contacts as well. In addition, the group enables the peer to join P2P e-commerce network and makes faking an identity more difficult.

Peers use self-certifying identifiers that are exchanged when they initially come into contact. These can be used as public keys to verify digital signatures on the messages sent by their neighbors. All communications between peers are digitally signed. In this kind of relationship, we use neighbor as our point of reference to address Sybil attack. Notice that in a group, whatever admission we set, there are honest, malicious, and Sybil attack peers who are authenticated by an admission control mechanism to join the group.

More honest peers are admitted compared to malicious peers, where the trust association is aimed at positive results. The knowledge of the social graph may reside in a single party, or be distributed across all users. In our work, we use the distributed admission control which only requires each peer to be initially aware of only its immediate trusted neighbors, and look for honest neighbors. The neighbors assist to locate

978-0-7695-4843-2/12 $26.00 © 2012 IEEE
DOI 10.1109/UIC-ATC.2012.20

other neighbors. We make an important observation about the challenges of Sybil resilient peers in admission. It has been impossible to get an algorithm which can detect all Sybil attack peers and identify all the honest peers. We further propose a centralized setting for admission control as long as the peers have already been partially admitted in a group.

In this paper, we present a distributed structured approach to Sybil attack. This is derived from the fact that our approach is based on the neighbor similarity trust relationship among the neighbor peers. Given a P2P e-commerce trust relationship based on interest, the transactions among peers are random as each peer can decide to trade with another peer any time. A peer doesn't have to consult others in the group unless a recommendation is needed. This approach shows the advantage in exploiting the similarity trust relationship among peers in which they are able to monitor each other.

Our contribution is threefold:

1) We propose *SybilTrust* that can identify and protect honest peers from Sybil attack. The Sybil attack identity peers will have their trust cancelled and dismissed from the group.
2) Based on the group infrastructure in P2P e-commerce, each neighbor is connected to the peers by the success of the transaction it makes or the trust evaluation level. A peer can only be recognized as a neighbor depending if its trust level is sustained over a threshold value.
3) *SybilTrust* enable neighbor peers carry recommendation identifiers among the peers in a group. This ensures the group detection algorithms to identify Sybil attack peer are efficient and scalable to large P2P e-commerce networks.

To achieve the results, *SybilTrust* uses a distributed algorithm to perform neighbor validation to ensure that the neighbor similarity trust information is kept as honest and secure as possible. This is able to limit the number of admitted Sybil attack peer identities to a very small number while admitting almost all the honest identities. After we admit a number of attack edges to cover more peers, the number of admitted Sybil attack peer identities remains very low. In this paper, we note that: 1) Sybil attack peers tend to be poorly connected to the rest of the network, compared to the non-Sybil peers; 2) Sybil attack peers use various graph analysis techniques to search for topological features resulting from the limited capacity of Sybil attack peer to establish social links.

Organization of this paper: First, we introduce related work in Section II. Next, we give system models and motivation in Section III. Section IV shows preliminaries. The proposed approach is presented in Section V. Section VI deals with trust evaluation between neighbor peers. Security and performance analysis is given in Section VII. Section VIII summarizes our conclusions and discusses future work.

II. Related Work

There has been much excitement in the research community using social networks to mitigate multiple identity, or Sybil attack. In the literature, Sybil attack can be thwarted by various ways: 1) binding identities to IP addresses, or IP prefixes [13] and 2) aaking every identity to solve puzzles that require human effort, such as CAPTCHAs [14]. The protection of the adversary can readily steal IP addresses with different prefixes in today's Internet [15].

Douceur is the first to describe Sybil attack in the context of P2P networks [1]. He proposed that Sybil attack can defeat redundancy mechanisms of distributed storage systems. He claimed trusted certification is the only approach that has the potential to completely eliminate Sybil attack. Trusted certification relies on a centralized authority that must ensure each entity is assigned exactly one identity, as indicated by possession of a certificate. Due to its single-point-failure problem, the Douceur approach may not be applicable to decentralized P2P e-commerce. In his initial paper on Sybil attack, he already proved a negative result showing that Sybil attack cannot be prevented unless special assumptions are made. He also proposed that, to prevent all Sybil attacks, the certifying authority must also ensure that no certificates are lost or stolen, which is probably impossible in almost all applications.

SybilGuard [5] uses a special kind of random walk, called random routes, in the social network. In this case, real-world social networks exhibit expander properties. The gatekeeper limits the number of admitted Sybil identities per attack edge. While its direction is promising, SybilGuard suffers from two major limitations. First, although the end guarantees of SybilGuard are stronger than previous decentralized approaches, they are still rather weak in the absolute sense. Each attack edge allows Sybil attack peers to be accepted. Second, SybilGuard critically relies on the assumption that social networks are fast-mixing, an assumption that had not been validated in the real world. SybilGuard assumes if there is a small set of attack edges, a number of nodes will be disconnected.

Ostra [16] assumes global knowledge about the social network. This method does not provide guarantees that are provable. Online communication media and social networking sites allow any sender to reach potentially millions of users at near zero marginal cost. Unfortunately, the same property opens the door to unwanted communications, marketing, and propaganda. An evaluation based on data gathered from an online social networking site shows that Ostra effectively thwarts unwanted communications while no impeding legitimate communications.

SumUp [11] uses adaptive maximum flow on the social network to bound the number of Sybil identities (voters) accepted per attack edge. It also assumes global knowledge about the social network. It leverages social network underlining the accounts to defend against Sybil attack. It mainly deals with online voting systems. The attacker prepares to cast bogus votes for an object. SumUp also leverages users' feedback to further reduce the number of bogus votes cast by the attacker, if the attacker keeps doing that for different objects.

SybilInfer [8] model proves that Sybil attack peers will increase the mixing time of the graph and thus affect the probability that a random walk starting from a region will

end within that region. This model alludes to decentralized designs, but none of them provides a complete design that is decentralized. There is no result proven on how much the probability is affected. Sybil-Infer determines the probability via sampling, which by itself has unknown estimation error. As a result, SybilInfer is not able to prove an end-to-end guarantee on the number of Sybil attack peers accepted. The algorithm proposed by Xu et al. [10] calculates the shortest path between every pair of nodes within the network in each round, which makes it impractical for even small-sized social networks. In contrast, SybilDefender only relies on performing a limited number of random walks in the social graph, and it scalable to large networks

Different from SybilLimit [3] with the goal of limiting the number of accepted Sybil nodes, Ostra and SumUp further leverage feedback to modify the weight of the edges in the social network dynamically.

Yu et al. [17] proposed DSybil, a novel defense for diminishing the influence of Sybil identities in recommendation systems. DSybil provides strong provable guarantees that hold even under the worst-case attack and are optimal. DSybil can defend against an unlimited number of sybil identities over time. DSybil uses feedback to defend against Sybil attack in the context of recommendation systems and provides strong provable end-to-end guarantees. In scenarios where feedback is available, we expect that combining these feedback-based techniques with neighbor similarity trust can further strengthen the defense. All the above, have not been able to give a guarantee to address Sybil attack.

A global trust model, EigenTrust [18], proposed by S. Kamvar is based on the transfer of trust. Iterating the trust of neighbor peers, EigenTrust calculates the global trust value built on a web of trust to reflect a peer's credibility. This method has a higher time complexity and a lower ability to resist risk. It adopts a binary rating function, interpreted as either positive one (representing satisfactory), zero, or negative one (representing unsatisfactory or a complaint). The work by Kamvar did not explicitly distinguish transaction reputation or recommendation reputation. EigenTrust does not take into account user dynamics, the effect of credibility, or attacks and threats.

Eigen group trust model in grouped P2P communities [19], proposed an effective trust system built on top of a P2P group infrastructure. The model is based on a delegation system that manages trust not only within communities, but also between different communities. It had its own weakness. First, it aimed at reducing the issue of overloading the network, which it never achieved satisfactorily. Second, it never took into consideration the issue of user dynamics, in which peers change their ways of operation, produce several identities, and become malicious in course of transactions. In their approach, how to effectively compute, or evaluate similarity degree between peers in a group is not investigated.

In contrast to all these efforts, our proposed *SybilTrust* uses the neighborhood similarity trust relationship to address Sybil attack. In the neighborhood, peers join groups of their own interest. The groups are formed by peers distributing information in a multipath by the small world phenomenon. In this case, the method ensures that the recommendation given is the right one to be delivered to the destination. We use the idea of identifiers in a distributed system so as to avoid the peers being compromised by Sybil and malicious peers. With combination of our method and the methods like SybilGuard and SybiLimit, aimed at using the feedback to address the Sybil attack, we can further strengthen the defense.

In summary, SybilGuard and SybilLimit have a number of design features that facilitate their use in decentralized systems. Similarly, SumUp [11] has optimizations specific to online content voting systems. Compared to other attacks, Sybil attack can easily interfere with many network protocols, such as voting, data aggregation, reputation evaluation, and so on. This is because a large portion of the nodes in the network could be illegitimate entities.

III. Models and Motivations

In this section, we describe our network model and the attack model.

A. Network Model

We consider a group with a number of peers which have open and anonymous characteristics. A peer can not make its own decisions on trust to another peer unless it is a member of the group. Each peer relates to other peers depending on the trust it has. A graph G is a tuple $\langle V, E \rangle$, where V is a set of vertices and E is a set of edges. Specifically, $V = \{v_1, v_2, \cdots, v_x\}$ represents the peers available, and $E = \{e_1, e_2, ..., e_y\}$ represents the edges among the peers. An edge is an ordered pair (v, z) of vertices, where v is called a trustor, and z is called a trustee. If vertex z is adjacent to vertex v, there is an edge (v, z) in E from v to z. Notice that if there is an edge (v, z) in E, then there is also an edge (z, v) in E.

The neighborhood of a peer v in a P2P e-commerce is $N(v) = \{z/(v, z) \in E\}$. Each peer v maintains a set of identifiers of its neighbors $N(v)$, in which each one is unique. Messages can be sent from a peer v to a peer z, provided that v knows the identifier of z. Any packet broadcast by a peer is received by all its neighbors. Each edge in E, for example, from peer a to peer b, has two trust factors, namely, trust value $t(a, b)$, and risk level $r(a, b)$, both of which take values from a real interval $(0, 1]$.

B. Attack Model

Some peers may launch arbitrary attacks to interfere with P2P e-commerce operations, or the normal functioning of the network. Major attacks in P2P e-commerce can be classified as passive and active attacks.

- *Passive attack*: It listens to incoming and outgoing messages, in order to infer the relevant information from the transmitted recommendations, i.e., eavesdroping, but doesn't harm the system. A peer can be in passive mode and later in active mode.

- *Active attack*: When a malicious peer receives a recommendation for forwarding, it can modify, or when requested to provide recommendations on another peer, it can inflate or bad mouth. The bad mouthing is a situation where a malicious peer may collude with other malicious peers to revange the honest peer.

In this paper, we focus on the active attacks in P2P e-commerce. When a peer is compromised, all the information will be extracted. In our work, we have proposed use of *SybilTrust* which is based on neighbor similarity relationship of the peers.

IV. Preliminaries

Our approach is different from the approaches already proposed, e.g., SybilLimit [3] and SybilGuard [4]. This approach is designed to work in a distributed setting where each peer is initially only aware of its immediate neighbors. The system consists of n peers who are either honest or malicious. The honest peers are more than the malicious peers for an ideal business scenario. There exists a unidirectional interest graph among all peers in a group. If two peers are honest, they represent an honest trust relationship. Each peer has six keys, namely, individual key, pairwise key, session key, group key, recommendation authentication code key (T_dAC), and Encryption key. Some keys in this approach may exist as private or public, depending on the magnitude and their individual applications. Mostly, when shared among all peers, they exist as public keys. They will also be private for the group members, as they cannot be shared to other member peers who have not been admitted to the group.

The Sybil attack peers collude with each other, and attack honest peers hence are referred to as adversaries. The Sybil attack peer can behave in Byzantine fashion. Our approach assumes an attacker has relationship with honest peers and may refer them as honest neighbors. Each peer acts as an admission controller, hence can judge which peer to admit and be its own controller to any malicious peer. The controller can be a suspect to others unless they have exchanged credentials and proved it is an honest peer. Each peer has a locally generated public/private key pair. The key generated by the relationship between two peers is known as a pairwise key.

Neighbor distribution for a particular interest to its neighbors can be a multi-hop link from a peer to its trusted neighbor, then to other neighbors. This is a small world phenomenon, where links created with distance neighborhood are very important in any business transaction. The trust being created increases as a peer links to others who validate it in different transactions.

In SybilGuard approach, a random walk starting from an honest peer in the social network is known as escaping, if it ever crosses any attack edge. Any connected social network with n peers, and g attack edges, has probability of length random walk starting from a uniformly random honest peer being escaping at most gl/n.

V. Our Proposed Approach

In this paper, our approach is in two parts, where part *A* deals with the detection of the attack and part *B* deals with distribution in neighbor similarity trust approach.

A. Similarity Trust Relationship

The SybilTrust protocol consists of two phases: a bootstrap phase, where each peer acts as an identifier source to disseminate identifier throughout the network, and a distribution phase, where each peer is determined whether it is a Sybil or not.

In our work, similarity of each pair of peers over the same set of neighbors is based on interest in a pair of peers, for instance $peer_i$ and $peer_j$, are represented as p_i, and p_j respectively. if N_i is the set of peer p_i's neighbors, and N_j is the set of peer p_j's neighbors. N_{ij} is the set of common neighbors of p_i and p_j assuming that the feedback is given by the peers which trade with that peer, hence which are in the same or different groups defined as S_{ij} is the similarity between p_i's and p_j's trust value, about the same set of neighbors. It can be defined by the feedback of p_i's and p_j's trust values, denoted as S_{ij}. If represents p_i's local feedback about p_j, this also shows p_i's behavior in different transactions. Considering the set of common neighbors of p_i, and p_j, we use $L(i,j)$ to present p_i's feedback about p_j, and the p_i's report about p_j's behavior, It is equal to as the trust value. Thus, $\overrightarrow{Q}_i =< L(i,H_1), L(i,H_2), ..., L(i,H_n) >$ is the p_i's trust vector about neighbors; $\overrightarrow{Q_j} =< L(j,H_1), L(j,H_2), ..., L(j,H_n) >$ is p_i's trust vector.

Similarity can be determined as the cosine angle between $\vec{Q}_i$ and $\vec{Q}_j$, whereby S_{ij} is calculated as equation (1):

$$S_{ij} = \frac{\sum_{x \in N_{ij}} (nL)_{ix} \times (nL)_{jx}}{\sqrt{\sum_{x \in N_{ij}} (nL)^2_{ix} \sum_{x \in N_{ij}} (nL)^2_{jx}}}, \quad (1)$$

if $\left\|\vec{Q}_i\right\|! = \left\|\vec{Q}_j\right\|! = 0$, and $S_{ij} = 0$, if $\left\|\vec{Q}_i\right\| = 1$, or $\left\|\vec{Q}_j\right\| = 0$. $[S_{ij}]$ denotes the matrix of neighbor similarity trust. We note that similarity relationship is symmetric [7], i.e, $S_{ij} = S_{ji}$.

B. Detection of Sybil Attack Based on Neighbor Similarity Trust

In Sybil attack, each malicious peer will forge multiple identity which does not physically exist within a network, in order to mislead the legitimate peers and honest peers into believing that they have many neighbors [4]. In this paper, we assume there are three kinds of peers in the system: legitimate peers, malicious peers, and Sybil attack peers. Each malicious peer cheats its neighbors by creating multiple identity, referred to as Sybil attack peers.

In this paper, P2P e-commerce communities are in several groups. A group can be either open or restrictive depending on the interest of the peers. We investigate the peers belonging to a certain interest group. In each group, there is a group leader who is responsible for managing coordination of activities in a group [20]. If the malicious peer is the leader of the group,

after several transactions can be discovered as per the feedback by neighbors and voted out. When peers join a group, they acquire different identities in reference to the group. Each peer has neighbors in the group and outside the group. Sybil attack peers forged by the same malicious peer have the same set of physical neighbors that a malicious peer has. Each neighbor is connected to the peers by the success of the transaction it makes or the trust evaluation level. To detect the Sybil attack, where a peer can have different identity, a peer is evaluated in reference to its trustworthiness, and the similarity to the neighbors. If the neighbors do not have same trust data as the concerned peer, including its position, it can be detected that the peer has multiple identity and is cheating. The method of detection of Sybil attack is depicted in Fig. 1. A_1 and A_2 refer to the same peer but with different identities.

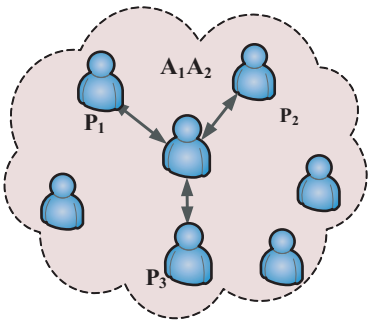

Fig. 1. Detection of Sybil attack.

When Sybil attack happens, A_1 and A_2 will both send messages.

$$\frac{M_{P1}^{A_1}}{M_{P2}^{A_1}} = \frac{M_{P1}^{A_2}}{M_{P2}^{A_2}}, \tag{2}$$

$$\frac{M_{P1}^{A_1}}{M_{P3}^{A_1}} = \frac{M_{P1}^{A_2}}{M_{P3}^{A_2}}. \tag{3}$$

If equations (2) and (3) are correct, Sybil attack must have happened, for the exclusive geographical position with two IDs. The most dangerous Sybil attack in P2P e-commerce is the outside intrusion. It means that the outside peers masquerade as an inside one to harm the network after catching the legitimate peers' key. The head peers communicate with the member peers in a group, and also other group heads. A peer communicates with a group leader occasionally. If the peer is just an ordinary member peer, it updates the leader every time. Member peer A_1, sends information to the group leader GL1 as shown in equation (4).

$$A_1 \rightarrow GL : \{ID_{A_1}, M(A_1)\}. \tag{4}$$

The GL compares the message with a message number to know whether the peer is honest or not by equation (5):

$$GL : \{|M(A_1) - M(A_2)| > X_M\}. \tag{5}$$

For an abnormal message, the peer detected is a Sybil attack peer. The GL leader occasionally releases flooding message to the group, where Sybil attack happened in peer A_1.

C. Distribution in Neighbor Similarity Trust Approach

In this section, we describe the distributed component of our *SybilTrust* and the challenges of the identifier distribution process.

In this paper, the principal building block of *SybilTrust* approach is the identifier distribution process. In the approach, all the peers with similar behavior in a group can be used as identifier source. They can send identifiers to others as the system regulates. If a peer sends less or more, the system can be a Sybil attack peer. The information can be broadcast to the rest of the peers in a group. We can use maximum flow computation as done in SumUp [11]. Any peer joining a group is assigned a unique identifier n_j, where j = 0, 1,$\cdots$, (N - 1), and N is the number of peers in the group. A peer has a peer identifier that is computed as in Chord [12], by hashing the IP address of the node. A peer a is a member of a group G defined as:

$a^n = a^n = aa\cdots a$; if $n > 0$ (n of a) or $a^n = e$; if $n = 0$ or $a^n = a^{-1}a^{-1}\cdots a^{-1}$; if $n < 0$ ($|n|$ of a^{-1}).

The order $|G|$ of a group G is its cardinality. A finite group whose order is a power of a prime p is called a p-group. In case there is another group in which the element is to power m, the rule holds as in (6):

$$a^m a^n = a^{m+n}, (a^m)^n = a^{mn}, \forall m, n \in Z. \tag{6}$$

From above, the group is:

$$\{n \in Z \,|a^n = e\}\ . \tag{7}$$

We assume that each peer x keeps $k = k(x)$ pointers to other peers. The peers are denoted as $l = \{l_1, l_2, ..., l_k\}$ where l_i is the distance between x and the i-th pointer. Without loss of generality, l is in a strictly ascending order, i.e., $l_1 < l_2 < ... < l_k$. When a request destined for peer y reaches peer x, then peer x will forward it to the next peer $x + l_i$, where $l_i \leq y - x \leq l_i + 1$. The peer-pair neighborhood (e.g., distance) between peer x and y is denoted as a function, (x, y). The distance satisfies the triangle inequality. That is, for any three peers x, y, z in the network, inequality $(x, y) \leq (x, z) + (z, y)$ holds. We can further derive that: $|(x, z) - (z, y)| \leq (x, y) \leq (x, z) + (z, y)$. The neighbor similarity protocol can be viewed in two ways:

1) Decentralized Identifier Distribution: Each peer acts as an identifier source. At the bootstrap phase all the peers which have similarity are determined by the neighbor peers and given the role of identifier distribution. In case a peer is a Sybil attack peer, it will try to send its own identifiers and will not be able to know the number given to others. The peers which are identified as Sybil attack peers are suspended from the group.

In our work, the number of identifiers to be disseminated t, is not a fixed parameter. The time is taken as a determinant for the dissemination of identifiers. We use certificates to ensure that the genuine identifiers can be known, others who send different signatures are malicious, and can be detected immediately. The signature chain represents a solution for detecting

double-spenders. Alternative mechanism may include secure transferable e-cash schemes [8] which allow a source peer to act as a ”bank” issuing e-coins as tickets. Each peer sends back the ticket to the peer which sends it as a proof that the peer received the signature. A peer which acts as a Sybil attack peer with many identities can be detected.

2) Prevent Maliciousness in Determining the Link Costs: Each peer in the P2P network relies on other peers to forward its requests, and in return is expected to forward the requests sent by other peers. A self-interested user might choose to free-load by refusing to forward requests, conserving local bandwidth and showing source to destination. Handling cheating in estimating link cost is a challenging task. In this paper, we propose a way in which it can be handled in P2P e-commerce. If the message sent from peer i to destination peer j is expected to be $q+$ and what is received from the receiver is q. We can calculate the cost effectiveness to determine cheating. This is gotten from the ratio of the two values which determine the cost.

VI. Trust Evaluation between Neighbor Peers

Trust depends on a subject's observation on the object and the third party recommendations. P2P e-commerce features need a trust evaluation mechanism without central peers where peers monitor each other. The openness enables malicious peers to take advantage and launch Sybil attack to the other honest peers. The subject obtains the trust value of objects according to both direct and indirect trust values. Peer i is subject, which not only makes direct assessment of object j, but also makes indirect evaluation of object j through peers h, k, l. The dotted circle in Fig. 2 represents the communication

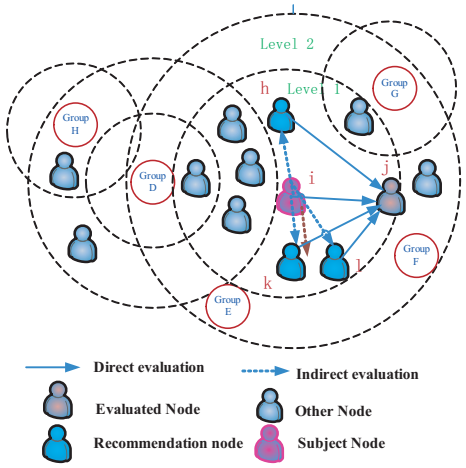

Fig. 2. The Recommendation Trust Relationship among Peers.

range of peer i and j respectively. This is from one level to another level. Peer i makes trust evaluation for peer j, and acknowledges by use of an acknowledgement mechanism. When the peer receives the recommendation, it sends back feedback information to the feedback source.

Our work assumes the intermediate peers are honest peers. The assumption made revokes the peer to broadcast the trust value it has. Depending on the assumptions a recommendation r, is received. If peer i makes a search on peer j, to confirm how many acknowledgements j sends as recommendations, the ratio of recommendations received by peer j can be obtained. We can detect whether peer j has a forging behavior. If the change maintains within $(-\lambda, \lambda)$ in different periods, peer j works normally. The calculation of r_{ij} given in (8) represents the received packets:

$$r_{ij}(t) = \frac{r_{ij}(t) - r_{ij}(t-1)}{r_{ij}(t) + r_{ij}(t-1)}, \tag{8}$$

if successful recommendations. $sr_{ij}\,(t)$ consist of j which sends the recommendations to k. Each particular recommendation has a time stamp. The equation (9) is:

$$sr_{ij}(t) = \frac{vr_{ij}(t)}{vr_{ij}(t) + wr_{ij}(t)}, \tag{9}$$

where $vr_{ij}\,(t)$ and $wr_{ij}\,(t)$ are the repeating recommendation. Each peer keeps one identifier to itself and distributes the rest evenly among its neighbors at the next level. In other words, a peer does not send tickets back to neighbors that are at the same, or smaller distance to the source. Since each peer only needs knowledge of its immediate neighbors to propagate identifiers. In our approach, the identifiers are only propagated by the peers who exhibit neighbor similarity trust.

Our perception is that, the attacker controls a number neighbor similarity peers, whereby a randomly chosen identifier source is relatively “far away” from most Sybil attack peer relationship. Every peer uses a “reversed” routing table. The source peer will always send some information to the peers which have neighbor similarity trust. However, if they do not reply, it can black list them. If they do reply and the source is overwhelmed by the overhead of such replies, then the adversary is effectively launching a DoS attack. Notice that the adversary can launch a DoS attack against the source. This enables two peers to propagate their public keys and IP addresses backward along the route to learn about the peers.

SybilTrust proposes that an honest peer should not have an excessive number of neighbors. The neighbors we refer should be member peers existing in a group. The restriction helps to bound the number of peers against any additional attack among the neighbors. If there are too many neighbors, *SybilTrust* will (internally) only use a subset of the peer's edges while ignoring all others.

Following Liben-Nowell and Kleinberg [6], we define the attributes of the given pair of peers as the intersection of the sets of similar products. Probability of the edge between $peer_i$ and $peer_j$ is $p_{pa}(i,j) = \alpha|C_i||C_j|$, where C_i is the set of products of:

$$AA(i,j) = \sum_{k \in C_i \cap C_j} \frac{1}{\log(|C_k|)}. \tag{10}$$

The function in equation (10) is zero when two peers share no products [7]. It creates a smooth distribution by interpolating

between the normalized Adamic-Adar score, and a preferential attachment model.

In a group each peer stores the trust data for the other member peers. A peer can be discovered to be malicious peer by determining the cost along the path when any information is send. The neighborhood of a vertex j is a set of vertices,

$$T_j = \{i : D(i,j) = 1\}. \tag{11}$$

For a given vertex in P2P e-commerce $j \in J$, let C_j be the local group coefficient of j, and it's equal to

$$C_j = |E(T_j)| / \begin{pmatrix} k_j \\ 2 \end{pmatrix}, \tag{12}$$

where $|E(T_j)|$ is the operator of counting the total number of links for all vertices in the set T_j. The group coefficient of a graph γ, denoted as $C(\gamma)$ in equation (13), is equal to

$$C(\gamma) = \frac{1}{N} \sum_{j \in J} C_j. \tag{13}$$

We consider a peer i and its neighbor peer j. N_i is the collection of the neighbor peers of i, while the neighborhood of peer i in the P2P e-commerce is $N(i) = \{j|(i,j) \in E\}$, where E represents the edge. We assume that each peer holds its own routing table, and on top of that it holds its neighbors routing tables. Thus, each peer has knowledge of a neighborhood of a given radius around it.

Let $G = (V, E)$ be a directed graph, where $V = \{v_1, v_2, ..., v_n\}$, and $l : (V \times V) \rightarrow S$ be a labeling function, where $(S, +, ., 0, 1)$ is a closed semi-ring. We take $l(v_i, v_j) = 0$, if $(v_i, v_j)$is not in E. For all i and j between 1 and n, the element $c(v_i, v_j)$ of S is equal to the sum over all paths v_i to v_j of the label path. We compute C_{ij}^k for all $1 \leq i \leq n, 1 \leq j \leq n$, and $0 \leq k \leq n$. Our aim is that C_{ij}^k should be the sum of the label paths from v_i to v_j such that all vertices on the path, except the end points, are in the set $\{v_1, v_2, v..., v_k\}$. The algorithm is as follows:

Algorithm: Computation Cost

1. **Input:** Graph $G = (V, E), v_i, v_j$, and the Trust
2. value $i, j, n, C(v_i, v_j)$
3. **Output:**$C(v_i, v_j)$
4. **For** $i \leftarrow 1$ **until** n **do** $C_{ii}^0 \leftarrow 1 + l(v_i, v_i)$;
5. **For** $1 \leq i, j \leq n$ and $i \neq j$ **do** $C_{ij}^0 \leftarrow l(v_i, v_j)$;
6. **For** $k \leftarrow 1$ **until** n **do**
7. **For** $1 \leq i, j \leq n$ **do**
8. $C_{ij}^k \leftarrow C_{ij}^{k-1} + C_{ik}^{k-1}.(C_{kk}^{k-1})^*.C_{kj}^{k-1}$
9. **For** $1 \leq i, j \leq n$ **do** $c(v_i, v_j) \leftarrow C_{ij}^n$
10. **end**

VII. Security and Performance Analysis

A. Security Analysis

We can illustrate the *SybilTrust* resilience by use of the controller in the peers to show that each controller only admitted the honest peers. Our method makes assumptions that the controller undergoes synchronization to prove whether the peers which acted as distributor of identifiers had similarity or not. If a peer never had similarity, the peer is assumed to have been a Sybil attack peer. Pairing method is used to generate an expander graph with expansion factor of high probability. Every pair of neighbor peers share a unique symmetric secret key (the edge key), established out of band [4] for authenticating each other.

A Sybil attack peer may disclose its edge key with some honest peer to another Sybil attack peer. However, because all neighbors are authenticated via the edge key, when A sends a message to B, B will still route the message as if it comes from B. In the protocol, every peer has a pre-computed random permutation (being the peer's degree) as its routing table. The routing table never changes unless the peer adds new neighbors, or deletes old neighbors. A random route entering via edge always exits via edge.

B. Performance Analysis

In this section, we evaluate the performance of the proposed *SybilTrust*. We measure two metrics, namely, non-trustworthy rate and detection rate. Non-trustworthy rate is the ratio of the number of honest peers which are erroneously marked as Sybil/malicious peer to the number of total honest peers. Detection rate is the proportion of detected Sybil/malicious peers to the total Sybil/malicious peers.

In our simulation, we use C++ tool. We ran an experiment consisting of 40 peers involved in 100 simulation runs resulting in a total of 4000 interactions.

Each honest and malicious peer interacted with a random number of peers defined by a uniform distribution. All the peers are restricted to the group. In our approach, P2P e-commerce community has a total of 40 different categories of interest. The transaction interactions between peers with similar interest can be defined as successful or unsuccessful, expressed as positive or negative respectively. The impact of the first two parameters on performance of the mechanism is evaluated. The percentage of malicious peers replied is randomly chosen by each malicious peer. Transactions with 10% to 40% malicious peers is done. Our *SybilTrust* approach detects more malicious peers compared to Eigen Trust [18] and Eigen Group Trust [19] as shown in Fig. 3.

Fig. 3. shows the detection rates of the P2P when the number of malicious peers increases. When the number of deployed peers is small, e.g., 40 peers, the chance that no peers are around a malicious peer is high. Fig. 3. illustrates the variation of non-trustworthy rates of different numbers of honest peers as the number of malicious peer increases. It is shown that the non-trustworthy rate increases as the number of honest peers and malicious peers increase. The reason is that when there are more malicious peers, the number of target groups is larger. Moreover, this is because neighbor relationship is used to categorize peers in the proposed approach. The number of target-groups also increases when the number of honest peers is higher.

As a result, the honest peers are examined more times, and the chance that an honest peer is erroneously determined as a Sybil/malicious peer increases, although more Sybil attack

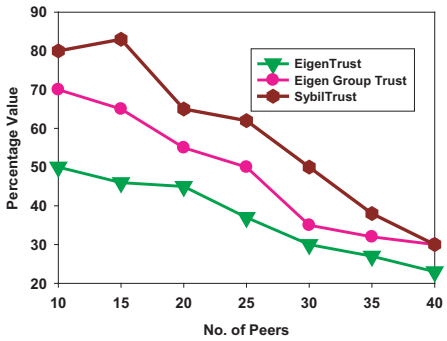

Fig. 3. Percentage of Peers that Detected the Malicious Peer.

peer can also be identified. Fig. 3. displays the detection rate when the reply rate of each malicious peer is the same. The detection rate does not decrease when the reply rate is more than 80%, because of the enhancement. The enhancement could still be found even when a malicious peer replies to almost all of its Sybil attack peer requests. Furthermore, the detection rate is higher as the number of malicious peers becomes more, which means the proposed mechanism is able to resist the Sybil attack from more malicious peers.

The detection rate is still more than 80% in the sparse network, which according to the definition of a sparse network is made in [19]. Moreover, the detection rate reaches 95% when the number of legitimate nodes is 300. It is also because the number of target groups increases as the number of malicious peers increases and the honest peers are examined more times. Therefore, the rate that an honest peer is erroneously identified as a Sybil/malicious peer also increases.

VIII. Conclusion and Future Work

We presented *SybilTrust*, a defense against Sybil attack in P2P e-commerce. Compared to other approaches, our approach is based on neighborhood similarity trust in a group P2P e-commerce community. This approach exploits the relationship between peers in a neighborhood setting. Our results on real-world P2P e-commerce confirmed fast-mixing property, hence validated the fundamental assumption behind SybilGuard's approach. For the future work, we intend to implement *SybilTrust* within the context of peers which exist in many groups. Neighbor similarity trust helps to weed out the Sybil attack peers and isolate maliciousness to specific Sybil peer groups rather than allow attack in honest groups with all honest peers.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under grant numbers 61073037 and 61103035, and the Ministry of Education Fund for Doctoral Disciplines in Higher Education under grant number 20110162110043.

References

[1] J. Douceur, "The Sybil Attack, " *Proc. of IPTPS*, 2002, pp. 251-260.
[2] A. Mohaisen, N. Hopper, and Y. Kim, "Keep Your Friends Close: Incorporating Trust into Social Network-based Sybil Defenses, " *Proc. of IEEE INFOCOM,* 2011, pp. 1-9.
[3] T. Nguyen, L. Jinyang, S. Lakshminarayanan, and S.M. Chow, "Optimal Sybil-Resilient Peer Admission Control," *Proc. of IEEE INFOCOM*, 2011, pp. 3218-3226.
[4] H. Yu, P. Gibbons, M. Kaminsky, and F. Xiao, "SybilLimit: A Near-Optimal Social Network Defense against Sybil Attack," *IEEE/ACM Transactions on Networking*, Vol. 18, No. 3, June 2010, pp. 3-17.
[5] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman, "SybilGuard: Defending against Sybil Attack via Social Networks," *IEEE/ACM Transactions on Networking*, Vol. 16, No. 3, June 2008, pp. 576-589.
[6] A. Tversky, "Features of Similarity, " *Psychological Review,* Vol. 84, No. 2, 1977, pp. 327-352.
[7] F. Musau, G. Wang, and M. B. Abdullahi, "Group Formation with Neighbor Similarity Trust in P2P E-Commerce," *Proceedings of the 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2011)*, pp. 835-840, November 16-18, 2011, Changsha, China.
[8] G. Danezis and P. Mittal, "SybilInfer: Detecting Sybil attack Peers using Social Networks," *Proc. of NDSS, San Diego, CA*, February 2009, pp.1-15.
[9] W. Wei, X. Fengyuan, C. T. Chiu, and L. Qun, "SybilDefender: Defend Against Sybil Attacks in Large Social Networks," *Proc. of IEEE INFOCOM,* 2012, pp.1951-1959.
[10] L. Xu, S. Chainan, H. Takizawa, and H. Kobayashi, "Resisting Sybil Attack by Social Network and Network Clustering," *International Symposium on Applications and the Internet IEEE/IPSJ SAINT*, 2010, pp.15 - 21
[11] N. Tran, B. Min, J. Li, and L. Subramanian, "Sybil-Resilient Online Content Voting," *Proc. of the 6th USENIX, Symposium on Networked Systems Design and Implement, USENIX Association*, 2009, pp. 15-28.
[12] I. Stoica, R. Morris, D. Karger, F. Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications," *Proc. of ACM SIGCOMM*, 2001, pp. 149-160.
[13] 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," *Proc. of ACM CCS*, 2002, pp. 207-216.
[14] L. V. Ahn, M. Blum, N. J. Hopper, and J. Langford, "CAPTCHA: Using Hard AI Problems for Security," *Proc. of IACR Eurocrypt*, 2003, pp. 294-311.
[15] A. Ramachandran and N. Feamster, "Understanding the Network-Level Behavior of Spammers," *Proc. of ACM, SIGCOMM*, 2006, pp. 291-302.
[16] A. Mislove, A. Post, K. Gummadi, and P. Druschel, "Ostra: Leveraging Trust to Thwart Unwanted Communication," *Proc. of USENIX NSDI*, 2008, pp. 15-30.
[17] H. Yu, C. Shi, M. Kaminsky, P. B. Gibbons, and F. Xiao, "DSybil: Optimal Sybil-Resistance for Recommendation Systems," *Proc. of Security Privacy Symbosium,* IEEE, 2009, pp. 283-298.
[18] S.D. Kamvar, M.T. Scholosser, and H.G. Molina , "The EigenTrust Algorithm for Reputation Management in P2P Networks," *Proc. of the 12th Int'l World Wide Web (WWW),* May 2003, pp. 640-651.
[19] A. Ravichandran and J. Yoon, "Trust Management with Delegation in Grouped Peer-to-Peer Communities," *SACMAT, ACM,* 2006 pp. 71-80.
[20] J. Wang, G. Yang, Y. Sun, and S. Chen, "Sybil Attack Detection Based on RSSI for Wireless Sensor Network," *Proc. of IEEE*, 2007, pp. 2684-2687.
[21] A. Mislove, M. Marcon, K. P Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Network," *Proc. of 7th ACM SIGCOM on Internet Measurement*, 2007, pp. 29-52.