Neighbor Similarity Trust against Sybil Attack in P2P E-Commerce

Guojun Wang, *Member, IEEE*, Felix Musau, Song Guo, *Senior Member, IEEE*, and Muhammad Bashir Abdullahi

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, an active attack, in which peers can have bogus and multiple identities to fake their owns. Most existing work, which concentrates on social networks and trusted certification, has not been able to prevent Sybil attack peers from doing transactions. Our work exploits the neighbor similarity trust relationship to address Sybil attack. In our approach, duplicated Sybil attack peers can be identified as the neighbor peers become acquainted and hence more trusted to each other. Security and performance analysis shows that Sybil attack can be minimized by our proposed neighbor similarity trust.

Index Terms—P2P, trust, Sybil attack, collusion attack, neighbor similarity

1 Introduction

P2P networks range from communication systems like e-mail 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. P2P overlay networks are known for their many desired attributes like openness, anonymity, decentralized nature, self-organization, scalability, and fault tolerance [18]. Each peer plays the dual role of client as well as server, meaning that each has its own control. All the resources utilized in the P2P infrastructure are contributed by the peers themselves unlike traditional methods where a central authority control is used.

Peers can collude and do all sorts of malicious activities in the open-access distributed systems. 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. The peer identities are then utilized to influence the behavior of the system. However, if a single defective entity can present multiple identities, it can control a substantial fraction of the system, thereby undermining the 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. To unify terminology, we call all identities created by malicious users as Sybil peers. In a P2P e-commerce application scenario, most of the trust considerations depend on the historical factors of the peers. The influence of Sybil identities can be reduced based on the historical behavior and recommendations from other peers. For example, a peer can give positive recommendations to a peer which is discovered is a Sybil or malicious peer. This can diminish the influence of Sybil identities hence reduce Sybil attack. A peer which has been giving dishonest recommendations will have its trust level reduced. In case it reaches a certain threshold level, the peer can be expelled from the group. Each peer has an identity, which is either honest or Sybil.

A Sybil identity can be an identity owned by a malicious user, or it can be a bribed/stolen identity, or it can be a fake identity obtained through a Sybil attack [24]. These Sybil attack peers are employed to target honest peers and hence subvert the system. In Sybil attack, a single malicious user creates a large number of peer identities called sybils. These sybils are used to launch security attacks, both at the application level and at the overlay level [18]. At the application level, sybils can target other honest peers while transacting with them, whereas at the overlay level, sybils can disrupt the services offered by the overlay layer like routing, data storage, lookup, etc. In trust systems, colluding Sybil peers may artificially increase a (malicious) peer's rating (e.g., eBay). Systems like Credence [3] rely on a trusted central authority to prevent maliciousness.

Defending against Sybil attack is quite a challenging task. A peer can pretend to be trusted with a hidden motive. The

- *G. Wang is with the School of Information Science and Engineering, Central South University, Changsha, 410083 Hunan Province, P. R. China. E-mail: csgjwang@csu.edu.cn.*
- *F. Musau is with School of Engineering and Technology, Kenyatta University, Nairobi City, 43844, 00100 Nairobi, Kenya. E-mail: musaunf@gmail.com.*
- *S. Guo is with School of Computer Science and Engineering, University of Aizu, Aizu-Wakamatsu City, Fukushima 965-8580, Japan. E-mail: sguo@u-aizu.ac.jp.*
- *M. B. Abdullahi is with Federal University of Technology, Minna, Nigeria. E-mail: el.bashir02@gmail.com.*

Manuscript received 2 Nov. 2013; revised 10 Mar. 2014; accepted 11 Mar. 2014. Date of publication 19 Mar. 2014; date of current version 6 Feb. 2015. Recommended for acceptance by J. Chen.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPDS.2014.2312932

peer can pollute the system with bogus information, which interferes with genuine business transactions and functioning of the systems [6]. This must be counter prevented to protect the honest peers. The link between an 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 attack is to use social networks to perform user admission control and limit the number of bogus identities admitted to a system [8], [9], [12], [14]. 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 the use of neighbor similarity trust in a group P2P e-commerce based on interest relationships, to eliminate maliciousness among the peers. This is refereed to as *SybilTrust*. In *SybilTrust*, the interest based group infrastructure peers have a neighbor similarity trust between each other, hence they are 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 avenue 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 a peer to join P2P e-commerce network and makes 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. We note that, all communications between peers are digitally signed. In this kind of relationship, we use neighbors as our point of reference to address Sybil attack. In a group, whatever admission we set, there are honest, malicious, and Sybil 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 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 to look for honest neighbors. The neighbors assist to locate other peers of same interest in other levels. 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 flexible as each peer can decide to trade with another peer any time. A peer doesn't have to consult others in a group unless a recommendation is needed. This approach shows the advantage in exploiting the similarity trust relationship among peers in which the peers are able to monitor each other.

Our contribution in this paper is threefold:

1) We propose *SybilTrust* that can identify and protect honest peers from Sybil attack. The Sybil peers can have their trust canceled and dismissed from a group.
2) Based on the group infrastructure in P2P e-commerce, each neighbor is connected to the peers by the success of the transactions it makes or the trust evaluation level. A peer can only be recognized as a neighbor depending on whether or not trust level is sustained over a threshold value.
3) *SybilTrust* enables neighbor peers to carry recommendation identifiers among the peers in a group. This ensures that the group detection algorithms to identify Sybil attack peers to be efficient and scalable in large P2P e-commerce networks.

To achieve these 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. *SybilTrust* is able to limit the number of admitted Sybil attack peer identities to a very small number while admitting most 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) the Sybil attack peers tend to be poorly connected to the rest of the network, compared to the honest peers, and 2) the Sybil attack peers use various graph analysis techniques to search for topological features resulting from their limited capacity to establish neighbor similarity links.

The rest of this paper is organized as follows. Section 2 introduces related work. Section 3 gives our system models and motivation. Section 4 shows preliminaries. The proposed approach is presented in Section 5. Section 6 deals with trust evaluation between neighbor peers. Security and performance analysis is given in Section 7. Section 8 concludes the paper.

2 Related Work

For a comprehensive discussion of the related work, please refer to Appendix A, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2014.2312932.

3 Models and Motivations

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

3.1 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 $|V| = n$ vertices and E is a set of edges. Specifically, $V = \{v_1, v_2, \ldots, v_x\}$ represents the peers available, and

$E = \{e_1, e_2, \ldots, e_y\}$ 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 $(0, 1]$.

Alternatively, we refer to $A = [a_{ij}]^{n\times n}$ as in [2] where the adjacency matrix $a_{ij} = 1$, if e_{ij} is in E and $a_{ij} = 0$. $P = [p_{ij}]^{n\times n}$ is the transition matrix

$$p_{ij} = \begin{cases} \frac{1}{d(v_i)} & e_{ij} \in E, \\ 0, & \text{otherwise}, \end{cases} \quad (1)$$

where $d(v_i)$ is the degree v_i, or the row norm of A:

$$d(v_i) = \sum_{k=1}^{n} a_{ik}. \quad (2)$$

The set of neighbors of v_i is $N(v_i)$ and $d(v_i) = |N(v_i)|$.

3.2 Attack Model

In order to launch a Sybil attack, a malicious peer must try to present multiple distinct identities. This can be achieved by either generating legal identities or by impersonating other normal peers. Some peers may launch arbitrary attacks to interfere with P2P e-commerce operations, or the normal functioning of the network. According to [4] an attack can succeed to launch a Sybil attack by:

- *Heterogeneous configuration*. in this case, malicious peers can have more communication and computation resources than the honest peers.
- *Message manipulation*. the attacker can eavesdrop on nearby communications with other parties. This means a attacker gets and interpolates information needed to impersonate others.

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., eavesdropping, 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 revenge the honest peer. In the Sybil attack, a malicious peer generates a large number of identities and uses them together to disrupt normal operation.

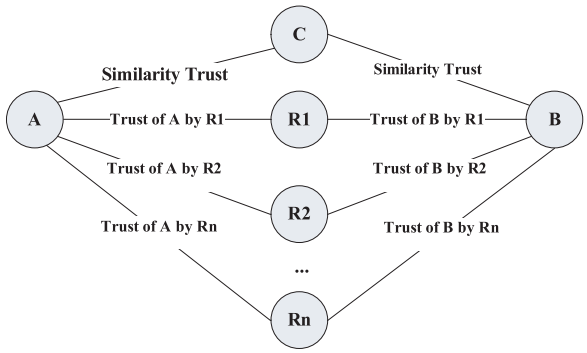


Fig. 1. Neighbor similarity computational model.

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. *SybilTrust* is efficient and scalable to group P2P e-commerce network.

4 Preliminaries

For a comprehensive discussion of the preliminaries, please refer to Appendix B, available online.

5 Our Proposed Approach

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

5.1 Neighbor Similarity Trust

In this section we present a Sybil identification algorithm that takes place in a neighbor similarity trust. The directed graph $G = (V, E)$ has edges and vertices. In our work, we assume V is the set of peers and E is the set of edges. The edges in a neighbor similarity have attack edges which are safeguarded from Sybil attacks. A peer u and a Sybil peer v can trade whether one is Sybil or not. Being in a group, comparison can be done to determine the number of peers which trade with peer.

If the peer trades with very few unsuccessful transactions, we can deduce the peer is a Sybil peer. This is supported by our approach which proposes peers existing in a group has six types of keys. The keys which exist mostly are pairwise keys supported by the group keys. We also note if an honest group has a link with another group which has Sybil peers, the Sybil group tend to have information which is not complete. Our algorithm adaptively tests the suspected peer while maintaining the neighbor similarity trust connection based on time.

5.1.1 Computational Model

In Fig. 1, if the recommendations given by the recommenders have a minimal difference, the peers are not Sybil peers. If the peer which has their similarity has trust which doesn't have a lot of variations, we can say that it's not a Sybil peer. Any peer who shows a lot of variation can be a Sybil peer hence classified as Sybil instead of honest peers.

In this approach, the attack edge is keenly monitored depending on the trust levels. The trust level can be interpreted as a probability; it can easily be integrated in decision making. Beyond simply choosing the best candidate available, the integration in utility-based decision making is possible. Any peer showing a lot of variation is a Sybil peer hence classified as Sybil instead of honest peer. In this approach the attack edge is keenly monitored depending on the Trust levels. The trust value can be interpreted as a probability; it can easily be integrated in decision making.

5.2 Threats from Compromised Peers

The Sybil attack peers may attempt to compromise the edges or the peers of the group P2P e-commerce. The Sybil attack peers can execute further malicious actions in the network. The threat being addressed is the identity active attacks as peers are continuously doing the transactions. Compromised peers may deliberately cause Byzantine faults in which their multiple identity and incorrect behavior ends up undetected. The Sybil attack peers can create more non-existent links. The protocols and services for P2P, such as routing protocols must operate efficiently regardless of the group size. In the neighbor similarity trust, peers must have a self-healing in order to recover automatically from any state. Sybil attack can defeat replication and fragmentation performed in distributed hash tables. Geographic routing in P2P can also be a routing mechanism which can be compromised by Sybil peers.

5.3 Cooperation among Peers in a Neighborhood

Cooperation is the strategy of a group of entities working together to achieve a common or individual goal [7]. Cooperation can be seen as an action of obtaining some advantage by giving, sharing, or allowing something. In cooperation we assume all the participants gain. In P2P e-commerce success will depend on a large measure of whether neighboring self-interested individuals have provided a structure, where proper incentives can act in a cooperative manner. All networking functions must be performed by the peers themselves where each peer acts as a router. The peers have to cooperate to communicate, discover, maintain the routes to other peers, and forward packets to their neighbors. In this cooperation some peers may gain advantage and propagate malicious transactions.

Among the peers, there are malicious and selfish peers which don't cooperate with others. In our research, we note the relationship between an evaluating peer and a peer being evaluated is worth exploring for similarity. It can help the reputation model decrease malicious evaluation, collect more subjective evaluations, and eventually calculate the global trust value. A neighborhood need to have incentives offered to the peers in order to encourage them to cooperate. In P2P we can classify incentive schemes into neighbor similarity-based system and payment-based system. Cooperation aims to reduce strategy peers which initially behave well and get high trust value after joining a network. Afterwards, they start to behave maliciously reducing QoS and providing dishonest feedback. The P2P neighbor similarity process must be a mutual trust level relation. Feedback evaluation among the peers is normally in accord with service evaluation, i.e., "will give you what you give me". Honest nodes provide honest service and feedback, while dishonest nodes provide neither honest service nor honest feedback whether they have a similarity relationship or not.

5.4 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 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. We consider the Jaccard metric whereby similarity is defined as follows:

$$sim(p_i, p_j) = \frac{|p_i \cap p_j|}{|p_i \cup p_j|}, \tag{3}$$

where $|p_i \cup p_j| \neq 0$. If $sim(p_i, p_j)$ is not smaller than the similarity threshold S, then the interests of p_i and p_j are similar. In the same logic presented, we can still determine the dissimilarity between peers which is not the scope of this paper. Therefore, dissimilarity between peers is

$$sim_\delta(p_i, p_j) = 1 - sim(p_i, p_j) = \frac{|p_i \cup p_j| - |p_i \cap p_j|}{|p_i \cup p_j|}. \tag{4}$$

We note that similarity relationship is symmetric [11], i.e, $sim(p_i, p_j) = sim(p_j, p_i)$. Similarity can be determined as the cosine angle between $\vec{Q}_i$ and $\vec{Q}_j$, whereby S_{ij} is calculated as:

$$S_{ij} = \frac{\sum_{x \in N_{ij}} (nL)_{ix} \times (nL)_{jx}}{\sqrt{\sum_{x \in N_{ij}} (nL)_{ix}^2 \sum_{x \in N_{ij}} (nL)_{jx}^2}}, \tag{5}$$

if $\|\vec{Q}_i\|! = \|\vec{Q}_j\|! = 0$, and $S_{ij} = 0$ otherwise. Let $[S_{ij}]$ denote the matrix of neighbor similarity trust.

5.5 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 [8]. In this paper, we assume there are three kinds of peers in the system: legitimate peers, malicious peers, and Sybil peers. Each malicious peer cheats its neighbors by creating multiple identity, referred to as Sybil 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 [27]. 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

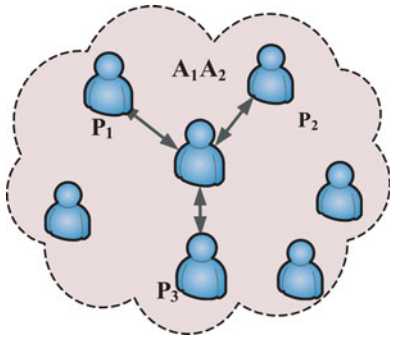


Fig. 2. Detection of Sybil attack.

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. 2. A_1 and A_2 refer to the same peer but with different identities.

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

$$\frac{M_{P1}^{A_1}}{M_{P2}^{A_1}} = \frac{M_{P1}^{A_2}}{M_{P2}^{A_2}}, \tag{6}$$

$$\frac{M_{P1}^{A_1}}{M_{P3}^{A_1}} = \frac{M_{P1}^{A_2}}{M_{P3}^{A_2}}. \tag{7}$$

If equations (6) and (7) 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 group leader peer 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 group leader every time. Member peer A_1, sends information to the group leader GL1 as shown in equation (8):

$$A_1 \rightarrow GL : \{ID_{A_1}, M(A_1)\}. \tag{8}$$

The GL compares the message with a message number to know whether the peer is honest or not by equation (9):

$$GL : \{|M(A_1) - M(A_2)| > X_M\}. \tag{9}$$

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.

5.6 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 having 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 [16]. Any peer joining a group is assigned a unique identifier n_j, where $j = 0, 1, \ldots, (N-1)$, and N is the number of peers in the group. A peer has a identifier that is computed as in Chord [17], by hashing the IP address of the node. A peer p 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 $n = 0$

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 that:

$$a^m a^n = a^{m+n}, (a^m)^n = a^{mn}, \forall m, n \in Z. \tag{10}$$

From above, the group is:

$$\{n \in Z | a^n = e\}. \tag{11}$$

We assume that each peer x keeps $k = k(x)$ pointers to other peers. The peers are denoted as $l = \{l_1, l_2, \ldots, l_k\}$ where l_i is the distance between x and ith pointer. Without loss of generality, l is in strictly ascending order, i.e., $l_1 < l_2 < \cdots < l_k$. 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 peer y is denoted as a function, (x, y). The distance satisfies the triangle inequality [23]. 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: $|(x, z) - (z, y)| \leq (x, y) \leq (x, z) + (z, y)$.

5.6.1 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. This makes the peer identified by the others not to coordinate the distribution of identifiers. The peers which are identified as Sybil 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 as in ref. [12], 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 the peer received the

signature. A peer which act as a Sybil attack peer with many identities can be detected.

If an attacker succeeds in capturing a sufficient number of peers, it could compromise a number of keys in the group. This will make the issue of generating a usable sybil identity trivial. It can be addressed by pairwise keys which are generated by two neighbor peers.

5.6.2 Key Validation

Key validation is divided into indirect and direct key validation.

- *Direct validation*. each peer challenges an identity using limited knowledge it possesses and makes a decision in depended of other peers. The peers may not reach a decision which is bonded to all the others.
- *Indirect validation*. Peers may collaborate in validating a peer. This is a decision which in our group scenario can be in a group.

This paper notes that if indirect validation is not done well it can lead to blackmail attacks. It provides a strong defense against Sybil attacks. To validate an identity, the verifier challenges the identity by requesting it to prove that it possesses one or more keys it claims to have [13]. For instance if there exists $\exists K_i, K_i \in \Omega(ID'), K_i \notin S$, and if a legitimate entity E in the P2P network knows K_i, then E can discover that ID is cheating by challenging ID using K_i. This involves indirect validation. During validation every peer challenge any other peer in the network to determine whether it is a Sybil peer or not. In our case, validation can be done by the peers which have a neighborhood similarity. During the validation the peers can estimate the time for the peers which may be planning to have malicious acts, this can be done by constantly monitoring the peer properties at a given time interval.

In our approach, we consider an attacker that performs breadth-first search for each identifier, until he finds the required keys. The number of times an attacker can find a usable identity is expressed as a probability. We consider the full validation where each identity is challenged by all the other peers in the group, so that we can prove the identity the peer claims to have.

We use the method in ref. [13] where each identity is challenged by a number of d nodes. To calculate the probability that ID is a usable sybil ID, we condition over t, the number of keys in $\Omega(ID')$ that are also in S, i.e, $t = card(\Omega(ID') \cap S)$, where $card(A)$ denotes the cardinality of the set A. $\Pr(t)$ passes validation with d verifiers:

$$pr(t) = \frac{\binom{n}{t}\binom{m-n}{k-t}}{\binom{m}{k}}. \tag{12}$$

In P2P, a peer validates an identity by use of the pairwise key between two neighbor peers. The adversary can compromise the entire link between peers to compute the pairwise key between the two identities or he will know nothing between the two identity and any other node. To evaluate the probability that at least l spaces are compromised given c compromised peers, we can get a direct measure of the difficult of a Sybil attack when a validation mechanism is present. Let S_i be the event where space i is compromised.

Ref. [13] has proved that given c compromised peers,

$$Pr(S_i) = \sum_{j=\lambda+1}^{c} \binom{c}{j}\left(\frac{l}{m}\right)^j\left(1-\frac{l}{m}\right)^{c-j}. \tag{13}$$

5.6.3 Prevent Maliciousness in Determining the Link Costs

Each peer in the P2P network relies on other peers to forward its requests, and in turn 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 peer surrounded by neighbors. 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 got from the ratio of the two values which determine the cost.

6 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 levels. 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. 3 represents the communication 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 (15) represents the received packets:

$$r_{ij}(t) = \frac{r_{ij}(t) - r_{ij}(t-1)}{r_{ij}(t) + r_{ij}(t-1)}, \tag{14}$$

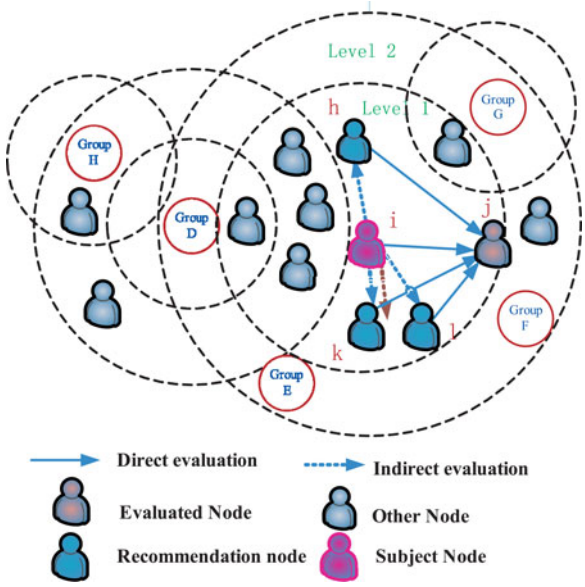


Fig. 3. The recommendation trust relationship among peers.

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

$$sr_{ij}(t) = \frac{vr_{ij}(t)}{vr_{ij}(t) + wr_{ij}(t)}, \tag{15}$$

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 of 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 [10], 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:

$$AA(i,j) = \sum_{k \in C_i \cap C_j} \frac{1}{\log(|C_k|)}. \tag{16}$$

The function is zero when two peers share no products [11]. It creates a smooth distribution by interpolating between the normalized Adamic-Adar score, and a preferential attachment model as shown in equation (16).

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,

$$T_j = \{i : D(i,j) = 1\}. \tag{17}$$

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

$$C_j = |E(T_j)| / \binom{k_j}{2}, \tag{18}$$

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 (19), is equal to

$$C(\gamma) = \frac{1}{N} \sum_{j \in J} C_j. \tag{19}$$

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 $V = \{v_1, v_2, \ldots, v_n\}$, and $l : (V \times V) \to S$ be a labeling function, where $(S, +, \ldots, 0, 1)$ 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 $1 \le i \le n, 1 \le j \le n$, and $0 \le k \le n$. 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 $\{v_1, v_2, v \ldots, v_k\}$. 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** $1 \le i, j \le n$ and $i \ne j$ **do** $C_{ij}^0 \leftarrow l(v_i, v_j)$;
6. **For** $k \leftarrow 1$ **until** n **do**
7. **For** $1 \le i, j \le n$ **do**
8. $C_{ij}^k \leftarrow C_{ij}^{k-1} + C_{ik}^{k-1}.(C_{kk}^{k-1})^*.C_{kj}^{k-1}$
9. **For** $1 \le i, j \le n$ **do** $c(v_i, v_j) \leftarrow C_{ij}^n$
10. **end**

6.1 Eliminating Sybil Communities

The Sybil attack detection problem can be addressed by finding an efficient algorithm to eliminate the Sybil groups which exist. The relationship between two peers by neighbor similarity trust is viewed as NP-complete. Assuming each neighbor relationship is a vertex in an undirected graph. We define the edge between the two vertices to be having a non-negative neighbor similarity value. Finding the groups is like finding all subgraphs, which is a well-known NP-complete problem. This can be represented as neighbor similarity of peer $p_i \cap p_j$.

$$S(p_i, p_j) = \begin{cases} -1 & +ve\,test \\ \frac{|p_i \cap p_j|}{\min\{|p_i|, |p_j|\}}, & -ve\,test, \end{cases} \tag{20}$$

where -1 represents that $p_i \cap p_j$ are distinct. $p_i \cap p_j$ denotes the set of common similarity peers. $|.|$ represents the size of a set or the length. We use neighbor similarity because the forged information issued from a malicious peer are similar (i.e., a non-negative similarity value) [4]. Communication provided by an honest peer may have connections with other communications which enable them to form a bigger group.

6.2 Accepting Honest Peers

Neighbor similarity trust introduce internal correlation within a single random route. Namely, if a random route visits the same peer more than once, the exiting edges will be correlated which is actually a feature in a group P2P e-commerce. In a neighbor similarity there is a small variation distance. Variation distance is a value in [0, 1] that describes the "distance" between two neighbor peers in a distribution. An honest peer should not have an excessive number of neighbor peers. This helps to identify Sybil peers. The random route protocol provides basic trust and security guarantees. SybilTrust ensures that all directed edge in the honest region allows only one public key to be registered. Honest peers have to be accepted based on their continuous trust level. In case the trust level drops other peers lose confidence with the peer. Moreover, two Sybil attackers will increase the resources required of the other in the trust levels by increasing participation, without increasing costs to honest participants. This brings the issue of participation as an issue for Sybil peers trying to outnumber the honest peers. Peers must prove themselves by offering benefits before getting anything in return, this means they must proof their trust level before they are engaged in any business transactions. Similarity trust introduce internal correlation within a single random route. Namely, if a random route visits the same peer more than once, the exiting edges will be correlated which is actually a feature in P2P e-commerce.

7 Security and Performance Analysis

7.1 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 [8] 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.

7.2 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.

Communication Cost. The trust level is sent with the recommendation feedback from one peer to another. If a peer is compromised, the information is broadcasted to all peers as revocation of the trust level is being done.

Computation Cost. The sybilTrust approach is efficient in the computation of polynomial evaluation. The calculation of the trust level evaluation is based on a pseudo-random function (PRF). PRF is a deterministic function.

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 4,000 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 percent malicious peers is done. Our *SybilTrust* approach detects more malicious peers compared to Eigen Trust [25] and Eigen Group Trust [26] as shown in Fig. 4.

Fig. 4. 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. 4 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

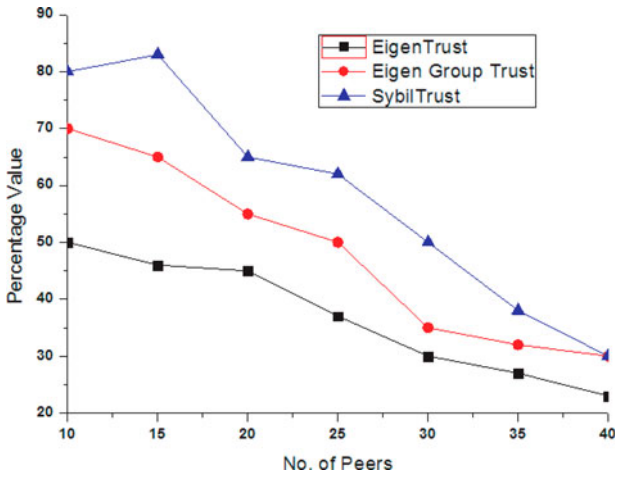


Fig. 4. Percentage of peers that detected the malicious peer.

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 peer can also be identified. Fig. 4 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 percent, 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 percent in the sparse network, which according to the definition of a sparse network is made in [26]. Moreover, the detection rate reaches 95 percent 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.

8 Conclusion

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. We also describe defense types such as key validation, distribution, and position verification. This methods can be done at in simultaneously with neighbor similarity trust which gives better defense mechanism. 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 peers and isolate maliciousness to specific Sybil peer groups rather than allow attack in honest groups with all honest peers.

Acknowledgments

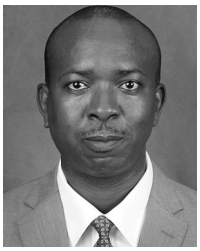
This work was supported by NSFC grants 61272151 and 61073037, ISTCP grant 2013DFB10070, the China Hunan Provincial Science & Technology Program under Grant Number 2012GK4106, and the Ministry of Education Fund for Doctoral Disciplines in Higher Education under Grant Number 20110162110043.

References

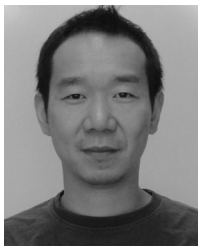
[1] J. Douceur, "The sybil attack," in *Proc. Revised Papers 1st Int. Workshop Peer-to-Peer Syst.*, 2002, pp. 251–260.
[2] A. Mohaisen, N. Hopper, and Y. Kim, "Keep your friends close: Incorporating trust into social network-based Sybil defenses," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2011, pp. 1–9.
[3] K. Walsh and E. G. Sirer, "Experience with an object reputation system for peer to peer filesharing," in *Proc. 3rd USENIX Conf. Netw. Syst. Des. Implementation*, 2006, vol. 3, pp. 1–14.
[4] S. Chang, Y. Qi, H. Zhu, J. Zhao, and X. Shen, "Footprint: Detecting Sybil attacks in urban vehicular networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 6, pp. 1103–1114, Jun. 2012.
[5] B. Yu, C. Z. Xu, and B. Xiao, "Detecting Sybil attacks in VANETs," *J. Parallel Distrib. Comput.*, vol. 73, no. 3, pp. 746–756, Jun. 2013.
[6] T. Nguyen, L. Jinyang, S. Lakshminarayanan, and S. M. Chow, "Optimal Sybil-resilient peer admission control," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2011, pp. 3218–3226.
[7] K. Wang, M. Wu, and S. Shen, "Secure trust-based cooperative communications in wireless multi-hop networks," in *Communications and Networking J. Peng, Ed.*, Rijeka, Croatia: InTech, Sep. 2010 ch. 18, pp. 360–378,.
[8] H. Yu, P. Gibbons, M. Kaminsky, and F. Xiao, "SybilLimit: A near-optimal social network defense against Sybil attack," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 3–17, Jun. 2010.
[9] H. Yu, M. Kaminsky, P.B. Gibbons, and A. Flaxman, "SybilGuard: Defending against Sybil attack via social networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 3, pp. 576–589, Jun. 2008.
[10] A. Tversky, "Features of similarity," *Psychological Rev.*, vol. 84, no. 2, pp. 327–352, 1977.
[11] F. Musau, G. Wang, and M. B. Abdullahi, "Group formation with neighbor similarity trust in P2P e-commerce," in *Proc. IEEE Joint Conf. Trust, Security Privacy Comput. Commun.*, Nov. 2011, pp. 835–840.
[12] G. Danezis and P. Mittal, "SybilInfer: Detecting Sybil attack peers using social networks," in *Proc. Netw. Distrib. Syst. Security Symp.*, San Diego, CA, USA, Feb. 2009, pp. 1–15.
[13] J. Newsome, E. Shi, D. Song, and A. Perrig, "The Sybil attack in sensor networks: Analysis and defenses," in *Proc. 3rd Int. Symp. Inf. Process. Sensor Netw.*, Apr. 2004, pp. 1–10.
[14] W. Wei, X. Fengyuan, C. T. Chiu, and L. Qun, "SybilDefender: Defend against Sybil attacks in large social networks," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2012, pp. 1951–1959.
[15] L. Xu, S. Chainan, H. Takizawa, and H. Kobayashi, "Resisting Sybil attack by social network and network clustering," in *Proc. Int. Symp. Appl.*, 2010, pp. 15–21.
[16] N. Tran, B. Min, J. Li, and L. Subramanian, "Sybil-resilient online content voting," in *Proc. 6th USENIX Symp. Netw. Syst. Des. Implementation*, 2009, pp. 15–28.
[17] I. Stoica, R. Morris, D. Karger, F. Kaashoek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup service for internet applications," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun*, 2001, pp. 149–160.
[18] B.S. Jyothi and D. Janakiram, "SyMon: A practical approach to defend large structured P2P systems against Sybil attack," *Peer-to-Peer Netw. Appl.*, vol. 4, pp. 289–308, 2011.
[19] 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 *Proc. 9th ACM Conf. Comput. Commun. Security*, 2002, pp. 207–216.
[20] L. V. Ahn, M. Blum, N. J. Hopper, and J. Langford, "CAPTCHA: Using hard AI problems for security," in *Proc. 22nd Int. Conf. Theory Appl. Cryptographic Tech*, 2003, pp. 294–311.

[21] A. Ramachandran and N. Feamster, "Understanding the network-level behavior of spammers," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun.*, 2006, pp. 291–302.
[22] A. Mislove, A. Post, K. Gummadi, and P. Druschel, "Ostra: Leveraging trust to thwart unwanted communication," in *Proc. 5th USENIX Symp. Netw. Syst. Des. Implementation*, 2008, pp. 15–30.
[23] P. Francis, S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavitt, and L. Zhang, "IDMaps: A global internet host distance estimation service," *IEEE/ACM Trans. Netw.*, vol. 9, no. 5, pp. 525–540, Oct. 2001.
[24] H. Yu, C. Shi, M. Kaminsky, P. B. Gibbons, and F. Xiao, "DSybil: Optimal Sybil-resistance for recommendation systems," in *Proc. IEEE Symp. Security Privacy*, 2009, pp. 283–298.
[25] S. D. Kamvar, M. T. Scholosser, and H. G. Molina, "The EigenTrust algorithm for reputation management in P2P networks," in *Proc. 12th Int. World Wide Web*, May 2003, pp. 640–651.
[26] A. Ravichandran and J. Yoon, "Trust management with delegation in grouped peer-to-peer communities," in *Proc. ACM Symp. Access Control Models Technol.*, 2006, pp. 71–80.
[27] J. Wang, G. Yang, Y. Sun, and S. Chen, "Sybil attack detection based on RSSI for wireless sensor network," in *Proc. IEEE on Wireless Commun., Netw. and Mobile Comput.* Sep. 2007, pp. 2684–2687.
[28] A. Mislove, M. Marcon, K. P Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social network," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas*, 2007, pp. 29–52.

Guojun Wang received the BSc degree in geophysics, MSc degree in computer science, and the PhD degree in computer science, from Central South University, China. He is a head and professor of the Department of Computer Science at Central South University. He is also a director of Trusted Computing Institute at Central South University. He has been an adjunct professor at Temple University; a visiting scholar at Florida Atlantic University; a visiting researcher at the University of Aizu, Japan; and a research fellow at the Hong Kong Polytechnic University. His research interests include network and information security, Internet of Things, and cloud computing. He is a distinguished member of CCF, and a member of the IEEE, ACM, and IEICE.



Felix Musau received BED (honors) degree in mathematics/computer science from Kenyatta University, Nairobi, Kenya, a postgraduate diploma in computer science and technology, MSC degree in computer science and technology and PhD in Computer Science from Central South University, China. He is a director of Information and Communication Technology at Kenyatta University, Nairobi, Kenya. He is also a lecturer at the Department of Computer Science at Kenyatta University. His current research interests include network and information security, trust management, E-Learning, and Internet of Things.



Song Guo received the PhD degree in computer science from University of Ottawa, Canada. He is currently a Full Professor at School of Computer Science and Engineering, the University of Aizu, Japan. His research interests are mainly in the areas of protocol design and performance analysis for computer and telecommunication networks. He received the Best Paper Awards at ACM IMCOM 2014, IEEE CSE 2011, and IEEE HPCC 2008. He currently serves as an Associate Editor of the *IEEE Transactions on Parallel and Distributed Systems*. He is in the editorial boards of *ACM/Springer Wireless Networks*, *Wireless Communications and Mobile Computing*, and many others. He has also served on organizing and technical committees of numerous international conferences, including as a general co-chair of MobiQuitous 2013. He is a senior member of the IEEE and the ACM.

Muhammad Bashir Abdullahi received the BTech (honors) degree in mathematics/computer science from the Federal University of Technology, Minna, Nigeria and the PhD degree in computer science from Central South University, China. He is a head of the Department of Computer Science at Federal University of Technology, Minna, Nigeria. His current research interests include trust, security and privacy issues in wireless sensor and ad hoc networks, Internet of things, and information and communication security.