# SW-R2P: A Trusted Small World Overlay P2P Network with Zero Knowledge Identification

Yingjie Xia
College of Computer Science, Zhejiang University, Hangzhou, P.R. China
xiayingjie@zju.edu.cn

Guanghua Song and Yao Zheng
College of Computer Science, Zhejiang University, Hangzhou, P.R. China
{ghsong, yao.zheng}@zju.edu.cn

***Abstract*—In order to implement both the efficiency and security in the Peer-to-Peer (P2P) network, we design a trusted small world overlay P2P network with the role based and reputation based access control policies, denoted as SW-R2P. The SW-R2P system integrates the small world topology with zero knowledge identification and Bayesian trust model. The zero knowledge identification is utilized to securely cluster all the peers into several groups without transferring any related information. The peer groups are then linked together to construct a trusted small world network based on the probabilities calculated by the Bayesian trust model. The simulation experiments demonstrate that the SW-R2P system achieves the performance with increased success rate in the resources lookup, strengthened robustness under the overwhelming traffic loadings, reduced reputation errors caused by the malicious peers and enhanced peer satisfaction rate for different trust metrics. In conclusion, the SW-R2P system collectively exploits the advantages of small world, zero knowledge identification and Bayesian trust model, therefore implementing a scalable, secure and efficient P2P network.**

***Index Terms*—Peer-to-Peer network, trusted small world, zero knowledge identification, Bayesian trust model, certificate**

## I. INTRODUCTION

The Peer-to-Peer (P2P) technology is a distributed information sharing solution with no centralized control. It has recently gained a lot of acceptance in many large scale distributed applications over the Internet, such as file sharing, streaming video delivery, multiple domain collaboration and distributed scientific computing. The most desired performance of the P2P applications is to locate resources and transfer data fast and efficiently, which can be further improved by constructing a small world paradigm over the P2P network [1].

The anonymity of the peers in the P2P network makes them vulnerable and unable to detect the attacks launched by selfish and malicious peers. However, if we use the exclusive identity to distinguish the peers and authenticate them during the data transfer, the peer identities can be easily embezzled by some attackers who then will masquerade to initiate harassment to the network. A zero knowledge identification scheme is adopted for all the peers to identify each other in a non-forgeable manner [2]. It can be implemented to serve as group identity authentication for constructing the small world topology of the P2P network.

In order to estimate the trustworthiness of the peers, the P2P network essentially involves a trust evaluation system. An efficient trust management strategy can encourage resources sharing among appropriate peers, and thus prevent buggy or malicious peers from their attacks. However, the unilateral trust models are not comprehensive enough to evaluate the multi-faceted trustworthiness of each peer. Therefore, the Bayesian trust model is proposed to combine the differentiated aspects of the trust. A Bayesian network is a relationship network that uses statistic methods to represent probability relationships between different elements [5]. It can be implemented in the P2P network to flexibly evaluate the different combinations of peer trust metrics, e.g. transfer quantity, resource quality.

By synthesizing the techniques introduced above, we design a P2P system called SW-R2P which denotes a trusted small world overlay P2P network with the role based and reputation based access control policies. These two policies are implemented by the zero knowledge identification and Bayesian trust model respectively. The contributions of the SW-R2P system are to implement the trusted small world P2P topology with zero knowledge identification and multi-faceted trust evaluation. With the advantages of all the involved techniques, the SW-R2P system achieves fast and secure resources lookup and data transfer, robustness against malicious peers, and high scalability in large scale P2P applications.

The remaining parts of this paper are organized as follows. Section 2 reviews related work on P2P networks with small world, zero knowldege interactive proof and

This paper is based on "R2P: A Peer-to-Peer Transfer System Based on Role and Reputation" which is appeared in proceedings of International Workshop on Knowledge Discovery and Data Mining, Australia, January, 2008. Guanghua Song is the corresponding author.

several trust models, and then briefly introduces our approach. In section 3 we specify the SW-R2P architecture. Then the detailed implementation and analysis are provided in section 4, and the simulated experiments and their analysis are presented in section 5. Finally, we draw a conclusion with a summary of contributions and propose some research plans.

## II. Related Work and Our Approach

The zero knowledge interactive proof (ZKIP) was first introduced by Goldwasser and Rackoff [3], and then gradually adopted in many authentication and identification protocols [6, 15]. The first application of ZKIP in P2P research is the PseudoTrust system, which supports the trust management in anonymous P2P networks [7]. However, other secret data in the P2P network, like group information, also need to be protected by ZKIP scheme, in order to avoid the embezzlement by malicious peers.

Abdule-Rahman and Hailes captured the most important properties of trust and reputation, and proposed a trust model and a recommendation protocol for distributed systems [8]. Wang and Vassileva introduced a Bayesian network based trust model in P2P networks, which provides a flexible method to represent differentiated trust in different aspects of each peer's capability [9]. However, these implementations only simply stay in the trust evaluation policies, but do not utilize the evaluated trustworthiness to construct the network topology.

The small world phenomenon was first observed by Milgram [10], where the six degrees of separation in a social network is described. Recent research work focuses on the small world theory applied in computer science. Hui, Lui and Yau implement a small world overlay P2P network, which provides a small world topological structure for the P2P network to efficiently look up and transfer resources [1]. However, this work only cares about the topology optimization, but neglects the important trust property.

Due to the shortages of these related works we implement the SW-R2P system which is a trusted small world overlay P2P network with zero knowledge identification and Bayesian trust model. The SW-R2P system evolves from R2P, a traditional structured P2P network with role and reputation based access control policies [4], and makes some improvements in robustness, scalability, security and efficiency. The zero knowledge identification in the SW-R2P system securely authenticates peers' group information to construct the peer groups which are the basic elements of the small world topology. The Bayesian trust model can provide the groups with different trust evaluations according to their distinct requirements, which promote the accuracy of the system behaviors. This trust model also supports linking the peer groups to construct a trusted small world topology. Finally we implement the Bayesian Decision Theory to select the most appropriate peers from the candidates. In summary, the SW-R2P system synthesizes the small world topology, zero knowledge identification and Bayesian trust model to implement an efficient, robust and secure P2P network. To the best of our knowledge, this system is the first work to synthesize the small world topology and trust model in P2P networks.

## III. The SW-R2P System Architecture

The architecture of the SW-R2P system is shown in Fig. 1. All the peers are categorized as the inner and exterior peers, which are clustered into several groups. There are also two types of links inside and between groups, namely group link and long link respectively. Each group has only one exterior peer, which takes at most $k$ long links to the other groups. The exterior peer also has the bidirectional group links to all the inner peers, which link to some other inner peers within the same group.

A peer who wants to join a group needs to use the zero knowledge interactive proof to verify its secret group information with the exterior peer of this group. The group information is assigned by the authority of the network, called trust center. Then the groups are long linked into the group network based on the interaction satisfactory probability calculated on the Bayesian trust evaluations. Finally the group network utilizes the Bayesian trust model again to create the trusted small world overlay P2P network. In this network each peer holds some trust instances to represent its multi-faceted trust evaluations to other peers which can support some system behaviors. All the details about the implementation of the SW-R2P system will be illuminated in the next section.

## IV. The SW-R2P System Implementation and Analysis

### A. Zero Knowledge Identification in SW-R2P

The zero knowledge identification in the SW-R2P system authenticates the peers' secret group information without transferring any related data. This will lead to securely constructing the small world topology. The whole process is composed of three phases: public values release, secret information preparatory and interactive proof procedure.

(a) Public Values Release

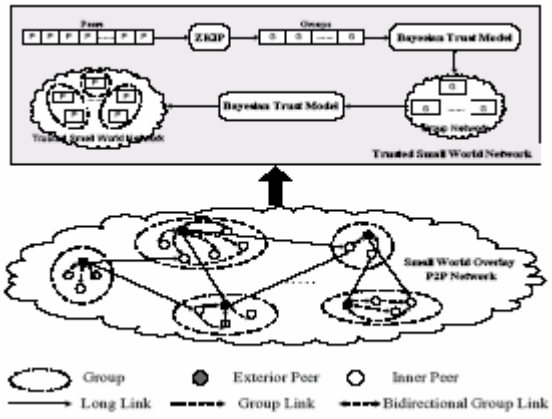



Figure 1. The SW-R2P system architecture.

The trust center releases the public value pair (*n*, *m*) group by group differently, and all the peers in the same group hold the same public values and can share them freely. In that value pair, the $n = p * q$ is a 1024-bit public number where *p* and *q* are two big primes. The other public number m is also determined to satisfy $m^k \equiv 1 \bmod n$ where *k* is a factor of the Euler's totient function $\phi(n)$.

(b) Secret Information Preparation

Group information $G_i$ is the secret information used to help peer *i* join the peers with the same group information as $G_i$. $n_i$ and $m_i$ are the two public values released by trust center for peer *i*. They prepare the secret group information $G_i$ and the certificate $C_i$ for peer *i*.

The trust center distributes a prime number $PK_i$ where $\gcd(PK_i, n_i) = 1$. Then the secret group information can be defined as $G_i \equiv m_i^{1/PK_i} \bmod n_i$. The certificate $C_i = (UID_i, PK_i, S_i)$ is the identity for peer *i*, where $UID_i$ is the only number for the peer and $S_i = signature(UID_i, PK_i)$ is generated by hash algorithm SHA-1, asymmetric encryption algorithm RSA and Base64 encoding algorithm [13]. After these preparations the peer *i* delivers the certificate $C_i$ to some other peers to verify whether it belongs to their groups or not.

(c) Interactive Proof Procedure

In this phase we denote the peer *i* as the prover and the target peer *j* as the verifier. The interactive proof procedure aims to check $G_i = G_j$ with transferring zero knowledge about the groups. The whole procedure is shown in Fig. 2.

The verifier peer *j* firstly checks the integrity and correctness of the prover's certificate $C_i = (UID_i, PK_i, S_i)$ by using the decryption algorithm reversing to the certificate generation. If the check result is valid, the $UID_i$ will be preserved into the routing table of peer *j* for the posterior resource lookups. Otherwise, the certificate is considered to be destroyed and peer *j* requires the peer *i* to send it again. After the check step, the prover sends the test string $x \equiv r^{PK_i} \bmod n_i$ to the verifier where the random number $r \in \{2, ..., n_i - 1\}$. As a challenge to the prover, the verifier sends a random number $e \in \{1, ..., 2^t\}$ where $2^t < PK_i$. And then, the prover peer *i* responses a witness $y \equiv rG_i^e \bmod n_i$. Finally, the verifier checks the equation $x \equiv y^{PK_j} m_j^{-e} \bmod n_j$. If the check result is equal, the prover peer *i* and the verifier peer *j* are proved to be in the same group. Otherwise the two groups are different and

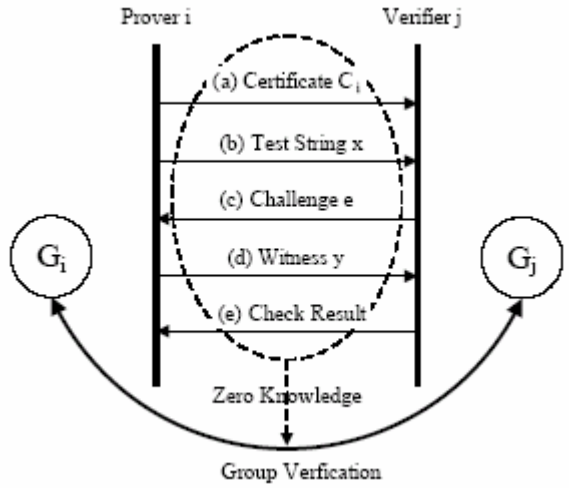



Figure 2. Zero knowledge identification procedure in SW-R2P.

the peer *i* needs to be verified with another group by the same procedure.

The completeness and soundness proofs of the zero knowledge identification in the SW-R2P system are provided for the validity and security verification.

***Lemma 1 (Completeness)*** If the prover peer *i* and the verifier peer *j* are in the same group, the zero knowledge check result should be always accepted as correct.

***Proof:*** If peer *i* and peer *j* are in the same group, thus $n_i = n_j$, $m_i = m_j$ and $PK_i = PK_j$. According to the definitions in phase (b) and (c), we can prove that

$$y^{PK_j} m_j^{-e} \bmod n_j = y^{PK_i} m_i^{-e} \bmod n_i = (rG_i^e)^{PK_i} m_i^{-e} \bmod n_i =$$
$$(r^{PK_i})(G_i^{PK_i})^e m_i^{-e} \bmod n_i = (r^{PK_i}) m_i^e m_i^{-e} \bmod n_i =$$
$$r^{PK_i} \bmod n_i = x. \quad (1)$$

***Lemma 2 (Soundness)*** Assume a malicious peer *p* does not know the secret group information of peer *j* $G_j$, and cannot compute the $PK_j$-th root in polynomial time. Following the zero knowledge identification, the check result between peer *j* and peer *p* will be unequal.

***Proof:*** The malicious peer *p* chooses a test string $x$ to compute two witnesses $y_1$ and $y_2$ from $y = \sqrt[PK_j]{x m_j^e \bmod n_j}$ on the challenges $e_1$ and $e_2$ respectively.

There exists two Bezout coefficients $q$ and $k$ for $PK_j * q + (e_1 - e_2) * k = \pm 1$ because $\gcd(PK_j, e_1 - e_2) = 1$. Therefore, according to the definitions in phase (b) and (c), we can educe that

$$(m_j^q (\frac{y_1}{y_2})^k)^{\pm 1} \bmod n = (G_j^{PK_j * q} (\frac{r * G_j^{e_1}}{r * G_j^{e_2}})^k)^{\pm 1} \bmod n =$$
$$(G_j^{PK_j * q + (e_1 - e_2) * k})^{\pm 1} \bmod n = G_j \bmod n. \quad (2)$$

Based on the formula $G_j = m_j^{1/PK_j}$ the revealed $G_j$ contradicts the assumption that the $PK_j$-th root can not be computed in polynomial time. Therefore, the peer *p* can not get the correct group information $G_j$ and pass the zero knowledge verification with peer *j*.

*B. Bayesian Trust Model in SW-R2P*

The Bayesian trust model is a Bayesian network based trust evaluation system, which can represent the multiple trust aspects. In the SW-R2P system we implement a simple Bayesian trust tree which is shown in Fig. 3. It is composed of one root node, overall trust ($OT$), and three leaf nodes, transfer quantity ($TQ$), resource quality ($RQ$) and download speed ($DS$).This trust model can be freely adjusted according to the requirements.

$TQ$ is a metric to denote the quantity contribution of the peer in the interactions. It is measured by utilizing the proportion of the upload and download data quantities to detect and avoid the free riders [11]. The resource quality $RQ$ is fed back to the original resource provider by the gainer. The download speed metric $DS$ is used to estimate the network condition of the provider. Both of them have five levels, which correspond to "2", "1", "0", "-1" and "-2" respectively. $OT$ is the root node of the Bayesian trust tree, which comprehensively evaluates the overall trust. Its value can be calculated by $V_{OT} = W_{TQ} * TQ + W_{RQ} * RQ + W_{DS} * DS$ where $W_{TQ} + W_{RQ} + W_{DS} = 1$. The $OT$ has two items "Satisfying" and "Unsatisfying", denoted by "1" and "0". By involving a trust threshold $\delta$, if $V_{OT} \geq \delta$, then $OT$ =1; otherwise $OT$ =0. $p(OT = 0 \mid 1) = (N_{unsat} \mid N_{sat}) / N$ represents that the probability to interact unsatisfactorily or satisfactorily can be measured in terms of the number of the corresponding interactions divided by the total number of the interactions.

Each peer holds the Bayesian trust instances of some other peers, which are mainly utilized in three components of the SW-R2P system.

(a) In a small world topology, the exterior peers need to create long links to some distant peers. The SW-R2P system calculates the long link probabilities based on the Bayesian trust instances whose different trust metrics are combined to satisfy the exterior peers' different requirements for long linked groups.

(b) During the P2P data transfer, the requester can combine several aspects of the provider candidates' trust models to calculate the probabilities which are used to select the most trustworthy peer as the resources provider.

(c) The role assignments and adjustments of the peers are determined by their evaluated trustworthiness. Their implementation details will be specified in the protocols of section 4.3.

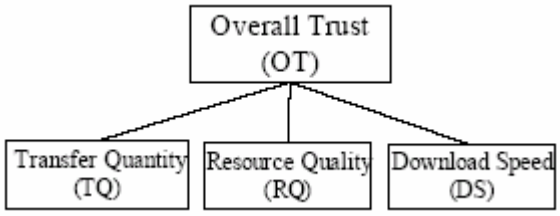



Figure 3. Implementation of Bayesian trust model in SW-R2P.

In the utilizations (a) and (b), the system needs to select one or some peers from the candidates according to their probabilities. Suppose $\{c_1, c_2, \ldots, c_n\}$ is the finite set of the candidates, and $\{\partial_1, \partial_2, \ldots, \partial_n\}$ is the corresponding set of the actions where $\partial_i$ denotes that the candidate $c_i$ is selected. The prior probability $P(c)$ is calculated on the Bayesian trust instances, to reflect how likely each candidate will be selected before we actually see it. If we simply select the candidate with the largest prior probability, the probability of the error for this decision will be $P(error)$ =min{ $1-P(c_1)$, $1-P(c_2)$, …, $1-P(c_n)$ }. This error estimation is not good enough, and thus we improve it based on the Bayesian probability.

Let $x$ be a continuous random variable which is the lightness measurement. $P(x \mid c_j)$ is defined as the class-conditional probability density representing the probability of $x$ given the candidate $c_j$. The SW-R2P system provides some hypothetical $P(x \mid c_j)$ functions for all the candidates. Furthermore, $\lambda(\partial_i \mid c_j)$ is defined as the error incurred for taking the action $\partial_i$ when the predicted candidate is $c_j$.

According to these definitions, the posterior probability can be computed by the Bayesian rule as $P(c_j \mid x) = P(x \mid c_j) * P(c_j) / P(x)$ where $P(x) = \sum_{j=1}^{n} P(x \mid c_j) * P(c_j)$. The expected error with taking action $\partial_i$ is $R(\partial_i \mid x) = \sum_{j=1}^{n} \lambda(\partial_i \mid c_j) * P(c_j \mid x)$ which also can be called the conditional risk. In order to simply analyze how minimizing the risk, we suppose that there are only two candidates to be selected. Therefore, the minimum-risk decision rule is that if $R(\partial_1 \mid x) < R(\partial_2 \mid x)$ we select the candidate $c_1$, otherwise we select $c_2$. Here we found that

$$R(\partial_1 \mid x) < R(\partial_2 \mid x)$$
$$\Rightarrow \lambda(\partial_1 \mid c_1) * P(c_1 \mid x) + \lambda(\partial_1 \mid c_2) * P(c_2 \mid x) < \lambda(\partial_2 \mid c_1) * P(c_1 \mid x) + \lambda(\partial_2 \mid c_2) * P(c_2 \mid x)$$

$$\Rightarrow (\lambda(\partial_2 \mid c_1) - \lambda(\partial_1 \mid c_1)) * P(c_1 \mid x) > (\lambda(\partial_1 \mid c_2) - \lambda(\partial_2 \mid c_2)) * P(c_2 \mid x)$$

$$\Rightarrow (\lambda(\partial_2 \mid c_1) - \lambda(\partial_1 \mid c_1)) * P(x \mid c_1) * P(c_1) / P(x) > (\lambda(\partial_1 \mid c_2) - \lambda(\partial_2 \mid c_2)) * P(x \mid c_2) * P(c_2) / P(x)$$

$$\Rightarrow \frac{P(x \mid c_1)}{P(x \mid c_2)} > \frac{\lambda(\partial_1 \mid c_2) - \lambda(\partial_2 \mid c_2)}{\lambda(\partial_2 \mid c_1) - \lambda(\partial_1 \mid c_1)} * \frac{P(c_2)}{P(c_1)}. \quad (3)$$

The condition is converted to compare the class-conditional probability ratio with a threshold that is independent of the measurement $x$. This decision algorithm based on the Bayesian probability can achieve a much better performance with the less decision errors than the simple prior probability based model.

### C. Trusted Small World Overlay P2P Network

The topology of the SW-R2P system is implemented by the trusted small world network, which adds the trust property to the common small world paradigm. All the peers are grouped by their secret group information, which is delivered by the trust center. The group information can be securely authenticated between peers by the zero knowledge identification. The only exterior peer of each group creates at most $k$ long links to the peers in other groups based on the probabilities calculated from the Bayesian trust model. As the interactions proceed, the exterior peer can adjust its long links based on the updated probabilities from the trust model.

The trusted small world overlay P2P network successfully synthesizes the trust and topology models, and can be mainly formed and maintained by four protocols for the behaviors of peer joining, peer leaving, peer failure and resource lookup.

(a) Peer Join Protocol (PJP)

If a peer $i$ needs to join a group $g$, we should first determine its role, inner or exterior. By reviewing all the interactions of peer $i$ and the exterior peer $j$ of $g$, we calculate the average trust probability per interaction $\overline{p}_{(Condition)} = (\sum_{k=1}^{n} p_{k\,(Condition)}) / n$, which estimates the likelihood to interact with the peer under some trust condition. If $\overline{p_i}$ is greater than $\overline{p_j}$, $i$ will join the group $g$ as an exterior peer, and move all $j$'s links to itself and create a bidirectional group link to the peer $j$. As a return all $i$'s original links from and to the inner peers of group $g$ are transferred to $j$. If peer $i$ joins $g$ as an inner peer, it will bidirectionally link to the peer $j$ and reserve all its original links from and to the inner peers of group $g$.

(b) Peer Leave Protocol (PLP)

Before peer $i$ leaves the group $g$, it has to find a substitute peer $j$ in the group. By the same algorithm to calculate the average trust probability in PJP, the peer $i$ determines the substitute with the smallest difference from $\overline{p_i}$, who is considered to hold the nearest trustworthiness to $i$.

If the group $g$ contains only one peer $i$, we can not find a substitute, and then the group will be removed. If the group $g$ contains more than one peer, then the substitute peer $j$ can be picked out, and the peer $i$ simply sends all its links to the peer $j$. $i$ also instructs all the peers long linking to itself to link to the substitute peer $j$. In the case that $i$ leaves the group $g$ and joins another group, the exterior peer of $g$ will create a long link to $i$.

(c) Peer Failure Protocol (PFP)

Periodically, each peer stores its interactions trust records in the trust center to tell that it is still alive. If a peer $i$ does not respond to other peer's request or it does not contact the trust center over the time limit, it is considered that $i$ has failed and the system loses all the information of $i$ except those trust records stored in the trust center. In that case, the system uses the peer failure protocol to recover and maintain some proper links.

Supposing that the peer $i$ is in the group $g$, if $g$ contains only one peer who is detected to have failed, $g$ will be removed from the system. If there are more than one peers in the group $g$, we have to find a substitute peer $j$ in group $g$ based on the $i$'s trust records stored in the trust center. The algorithm to find the peer $j$ is the same as to find a substitute in the PLP. When the substitute is determined, the system will instruct all the peers who have links to peer $i$ to reconnect to the substitute peer $j$. The peer $j$ also needs to inherits all the links of the peer $i$.

(d) Resource Lookup Protocol (RLP)

The resource lookup protocol is responsible for locating the required data object in the SW-R2P system. Supposing that the peer $i$ looks for the data $d$, it firstly asks the peers in its group who link to $i$ directly. If any of them holds $d$, the lookup process terminates successfully. Otherwise, we should check which role the peer $i$ takes. If $i$ is an exterior peer, it sends the lookup request to the long linked peers in other groups. If one of those peers has $d$, it responds with "yes" and the lookup process also succeeds. Otherwise, if all of them respond with "no", then those peers will use the RLP to look up recursively. If $i$ is an inner peer, it firstly group links to the exterior peer of its group, and then continues the resource lookup as the exterior peer described above.

Generally, there are many copies of data distributed in the network, and it is therefore likely that more than one peers respond to have the data. We need to select the most trustworthy one as the resource provider based on their trust evaluations.

## V. Simulation Experiments and Analysis

### A. Simulation Setup

The performance of the SW-R2P system is demonstrated by four simulated experiments: resource

lookup traversal, clustering coefficient, average reputation error and peer satisfaction rate, which will be deployed and run on two hundred personal computers in the public computing grid [14] of our university. All the computers are set up with Xeon Dual Core 2.4G CPU and 1GB memory. They are connected by the Ethernet of 1Gbps backbone and 100Mbps to desktop. Theoretically we can simulate peers on each computer as many as possible by using different network ports, unless the computer performance drops sharply due to overloading. Therefore, it is preferable for all the simulated peers to be distributed on the computers evenly.

We compare the SW-R2P system with the Chord and R2P systems in these experiments. We also configure and deploy both the systems in the simulated P2P network. Chord is a scalable P2P framework which can provide an efficient method to locate the resources [12]. And the R2P system is a structured P2P network with an excellent unilateral trust model [4]. In experiments A and B, we compare SW-R2P with Chord and R2P to measure their performances in resource lookup under different network conditions. And in experiments C and D, due to no trust evaluation strategy in Chord, only the trust models of the SW-R2P and R2P systems are evaluated in terms of their robustness, accuracy and flexibility.

The Bayesian trust model of the SW-R2P system in these simulation experiments are simply implemented as a tree with three nodes. The root node is still the overall trust $OT$, and the other two leaf nodes are the transfer quantity $TQ$ and resource quality $RQ$. Since the trust evaluation model of the R2P system unilaterally calculates the sum of the weighted $TQ$ and $RQ$ [4], it is appropriate to compare the SW-R2P system with the R2P system on their trust models.

The simulation parameters and their default values are summarized in Table 1. The initial network of the SW-R2P system consists of 1000 simulated peers in total, averagely 5 peers on each personal computer. Each group is restricted to contain no more than 100 peers, and the exterior peer has a maximum long link number which is initially set to 4. At the beginning, we also assume 98% of the peers are honest, and 2% are malicious which aim to disturbing the SW-R2P network by various behaviors. The default proportions of the peers to require different trust metrics are set [1/3, 1/3, 1/3]. All the experiments take $t$ =10 runs to calculate the average results.

TABLE I.
PARAMETERS AND DEFAULT VALUES IN SIMULATED EXPERIMENTS.

| Parameter | Basic definition | Default value |
|---|---|---|
| $num$ | Number of simulated peers | 1000 |
| $G$ | Maximum group size | 100 |
| $k$ | Maximum long links | 4 |
| $p$ | Malicious peer percentage | 2% |
| $P$ | Peer proportions to require different trust metrics | [1/3, 1/3, 1/3] |
| $t$ | Run times of experiments | 10 |

*B. Experiments and Analysis*

**Experiment A:** Performance of random resources lookup under various system sizes and long link limits.

This experiment is conducted to measure the average number of lookup hops between two peers. We simulate the network with the size $num$ =1000, 2000, …, 10000. One peer initially holds an exclusive resource object, which means the network has $num$ distinct objects in total. Each peer looks up a randomly chosen object for $t$ times in order to calculate the average link traversals, which can be aggregated for the whole network. Parameter $k$, the maximum long links of the exterior peer in the SW-R2P system, is set as 4, 8 and 12 respectively in this experiment. Fig. 4 illustrates the average link traversals of the random resources lookup in the SW-R2P, Chord and R2P systems with various network sizes and long link limits.

From Fig. 4, we observe that from $num$ =1000 to 10000 the average link traversals for all the SW-R2P systems are less than 7, and their increments also do not exceed 1. However, the compared Chord and R2P systems rise sharply from about 6 to almost 10. It indicates that our SW-R2P system not only has lower link traversals, but also keeps them stable under various network sizes. As the long link limit increases from 4 to 8 and 12, the values of the SW-R2P system decrease from around 6 to 5 and 4. This is because the more neighbors of each group shorten its average distances to other groups. These indicated advantages of the SW-R2P system are mainly brought about by its small world topology.

**Experiment B:** Clustering coefficient under various network sizes.

The clustering coefficient measures the compactness degree of the P2P network, which shows the system robustness to look up resources under overwhelming traffic loadings [1]. It can be defined as (4) where $V$ is the peer set; $\Gamma_v$ is the neighbor peer set of $v$; $E(\Gamma_v)$ denotes the total number of links for all the peers in $\Gamma_v$ and $k_v$ is the size of the set $\Gamma_v$.

$$C = \frac{1}{num}\sum_{v \in V}\left(E(\Gamma_v) / \binom{k_v}{2}\right) \tag{4}$$

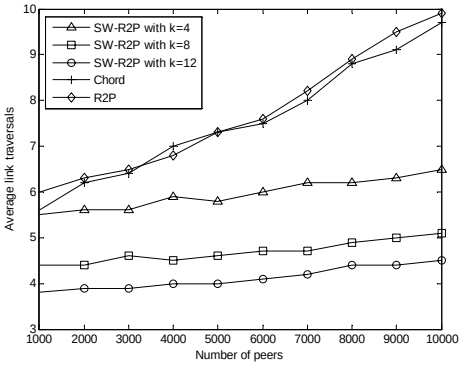



Figure 4. Average link traversals of random resources lookup in the SW-R2P, Chord and R2P systems.

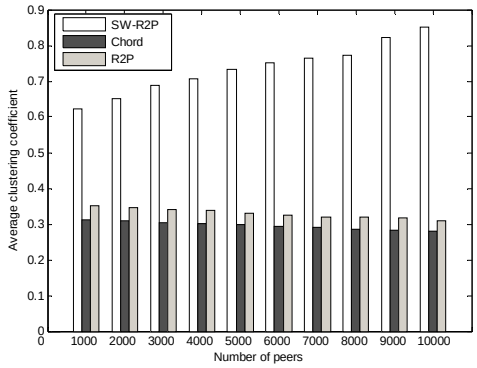



Figure 5. Average clustering coefficients of the SW-R2P, Chord and R2P systems.

By setting $num$ =1000, 2000, …, 10000 and $k$ =8, the average clustering coefficients of the SW-R2P, Chord and R2P systems are shown in Fig. 5. We observe that the SW-R2P system has more than twice the clustering coefficients as the Chord and R2P systems. As the network size increases, the clustering coefficient value of the SW-R2P system goes up from about 0.6 to 0.85 due to the group structure containing more peers in larger network. However, the Chord and R2P systems decrease their clustering coefficients gradually because their peers have few clustering relationships. Since the R2P system has a larger DHT size than the Chord system, its values are correspondingly a bit higher.

**Experiment C:** Average reputation error.

In this experiment we evaluate the security of the SW-R2P system against various malicious behaviors which can be categorized below.

(a) Masquerade another peer's identity to leak the secret resources, destroy its reputation, etc.

(b) Evaluate the peers who provide good services very low and those who provide bad services very high.

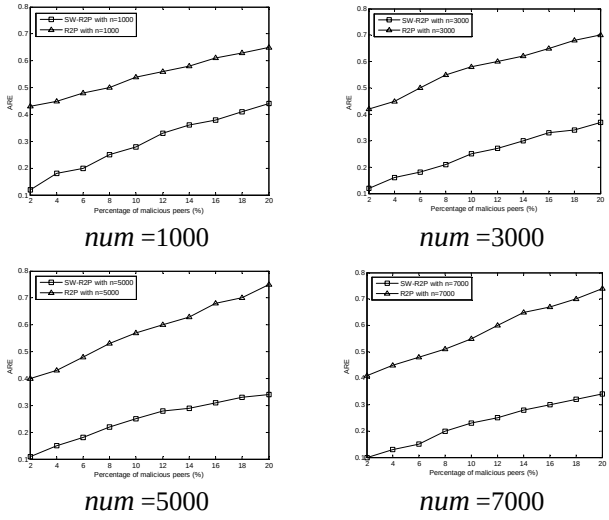
(c) Collaborate with each other to intentionally boost up their reputations.

We define the average reputation error (ARE) to evaluate the reputation error caused by the malicious peers in (5). A lower ARE indicates a higher immunity to the malevolent behaviors, and vice versa.

$$ARE = \frac{1}{num} \sum_{v \in V} \frac{|T_v - T'_v|}{T_v}. \qquad (5)$$

The $T_v$ and $T'_v$ are actual and measured overall trust scores $OT$ of peer $v$ respectively.

In this experiment we set the number of the malicious peers from 2% to 20% of total $num$ =1000, 3000, 5000 and 7000 peers, which are divided into the three categories evenly. The ARE values of the SW-R2P and R2P systems under various $p$ and $num$ conditions are plotted in Fig. 6. Both of them increase along with the malicious peer percentage rising up, and the SW-R2P system always has much less ARE values than the R2P system, with the differences always kept more than 0.3. Furthermore, we also discover that the larger the network size is, the greater the ARE difference becomes. When $p$ equals to 20%, the differences for the network size




$num$ =1000 $num$ =3000

$num$ =5000 $num$ =7000

Figure 6. Average reputation errors (ARE) of the SW-R2P and R2P systems.

$num$ =1000, 3000, 5000 and 7000 increase from 0.2 to 0.4. It can be explained by that the larger systems bring more malicious peers and behaviors, which can be better handled by the more excellent security and trust strategies of the SW-R2P system. The zero knowledge identification scheme can eliminate the malicious behavior (a), and the Bayesian trust model which provides differentiated trust evaluations for the peers can reduce the harassments initiated by (b) and (c). Therefore, this experiment indicates that the SW-R2P system is more resistant to various abuses caused by malicious behaviors.

**Experiment D:** Satisfaction rate of all the peers.

This experiment is designed to evaluate the accuracy and flexibility of the Bayesian trust model in the SW-R2P system. We define a measurement named peer satisfaction rate (PSR) in (6) to denote the percentage of the satisfactory requests in all the resource requests. If a request can locate the most trustworthy provider according to its trust metric requirement, it will be considered as a satisfactory request.

$$PSR = (\sum_{k \in M} N_{k\&Sat}) / N. \qquad (6)$$

The $N$ is the number of all the resource requests, $M$ is the set of trust metrics and $N_{k\&Sat}$ denotes the number of satisfactory requests with $k$ metric.

We simulate the network with the size $num$ =1000, 2000, …, 10000. Each peer randomly holds one of the 100 different resources and looks up another random resource under the hop limit of 10. In the both SW-R2P and R2P systems, the relationships of the three trust metrics are simulated in three cases: $OT = 0.5*TQ + 0.5*RQ$, $OT = 0.8*TQ + 0.2*RQ$ and $OT = 0.2*TQ + 0.8*RQ$. All of them provide different weights for the $TQ$ and $RQ$ to calculate the $OT$. $P$, the proportions of the peers to require different trust metrics [$OT$, $TQ$, $RQ$] are set as [1/3, 1/3, 1/3], [1, 0, 0], [0, 1, 0] and [0, 0, 1]. Fig. 7 shows the PSR values of

the SW-R2P and R2P systems with the size $num$ =1000, 2000, …, 10000, different metrics relationships and different peer proportions.

In Fig. 7, the four charts in each row take the same metrics relationship, and the three charts in each column have the same proportions of the peers. In all these charts, as the network size increases the PSR values of the both systems decrease gradually because more peers bring a greater probability to exceed the hop limit during the resources lookup. We can also observe that all the PSR values of the SW-R2P system are more than 0.9. This indicates the SW-R2P system can accurately evaluate the multi-faceted trusts which are required by different peers. No matter what the weights of $TQ$ and $RQ$ are, by setting $P$ =[1/3, 1/3, 1/3] in the charts (a), (e) and (i), the PSR values of the R2P system are always less than 0.52 and the differences between the SW-R2P and R2P systems keep around 0.5. If $P$ is set as [1, 0, 0] in (b), (f) and (j), the PSR values of the both systems are similar, because all the peers only consider the $OT$ trust metric evaluated in both SW-R2P and R2P systems. In the chart (c), when $P$ =[0, 1, 0] and $OT = 0.5*TQ + 0.5*RQ$, the PSR values of the R2P are less than 0.35 and the differences between the SW-R2P and R2P are around 0.6. This is similar to the chart (d) whose proportion is equal to [0, 0, 1]. However, if $OT = 0.8*TQ + 0.2*RQ$, the PSR values of R2P in chart (g) are about 0.6 which are much more than the chart (c), while its values in chart (h) are around 0.2 which are much less than the chart (d). These are all caused by the peer proportions to regard different trust metrics coordinated with the metric weights in the relationship. Due to the same reason, under $OT = 0.2*TQ + 0.8*RQ$ the lines of the R2P system in (k) and (l) are plotted similarly to (h) and (g) respectively. This experiment proves that the SW-R2P system can provide an accurate and flexible trust evaluation under the various combinations of trust metric relationships and peer proportions.

## VI. Conclusions and Future Work

In this paper, we introduce the design, implementation and simulated performance of the SW-R2P system, a trusted small world overlay P2P network with zero knowledge identification scheme and Bayesian trust model. Our contributions are summarized in three aspects.

(a) Trusted small world topology. Through the trust evaluation in the links creation and the protocols of the small world topology construction and maintenance, the SW-R2P system can provide a scalable, robust and secure performance in resources lookup and data transfer. As far as we know, the trusted small world topology of the SW-R2P system is the first work to integrate the small world paradigm with a trust management strategy in P2P networks.

(b) Identification with zero knowledge. As the basic elements of small world topology, the peer groups in the SW-R2P system are organized by authenticating the peers' group information without transferring any related data. The zero knowledge identification eliminates the malicious behavior of the information embezzlement, and keeps the whole network secure.

(c) Multi-faceted trust evaluation. The SW-R2P system uses the Bayesian trust model to provide the trust evaluation in multiple aspects. Although there are some existed P2P systems employing the Bayesian trust network, the SW-R2P system extends its influence on the network topology.

For the future work, we suggest the following research tasks to be completed.

(a) Influence between topology and trust. Various methods to construct small world topology should be investigated, especially based on some trust models. Furthermore, we plan to do some research on the bidirectional influences between the topology and trust.

(b) Bayesian trust model. Another future task is to add more aspects to the Bayesian trust model, adjust their corresponding weights and build a more realistic relationship network for them. This work needs a large amount of test analysis and debugging.

(c) Applications. We plan to apply the SW-R2P system to some specific P2P applications, such as scientific computing, collaboration and etc. In particular, we are encouraged to use our structure to solve the copyright protection problem in the P2P networks for content delivery.

## Acknowledgment

This work was fully supported by the National Natural Science Foundation of China under grant Number 90412014. We appreciate the valuable discussions and suggestions from the colleagues in the Center for Engineering and Scientific Computation, Zhejiang University.

## References

[1] K. Hui, J. Lui, and D. Yau, "Small World Overlay P2P Networks: Construction, Management and Handling of Dynamic Flash Crowds", *Computer Network*, vol. 50, pp. 2727-2746, 2006.

[2] L. Babai, "Arthur-merlin Games: A Randomized Proof System, and a Hierarchy of Complexity classes", *Journal of Computer and System Sciences*, vol. 36, pp. 254-276, 1988.

[3] S. Goldwasser, and C. Rackoff, "Knowledge Complexity of Interactive Proof Systems", *SIAM Journal on Computing*, vol. 18, pp. 186-208, 1989.

[4] Y. Xia, G. Song, Y. Zheng, and M. Zhu, "R2P: A Peer-to-Peer Transfer System Based on Role and Reputation", *Proc. of International Workshop on Knowledge Discovery and Data Mining*, Adelaide, Australia, 2008, pp. 136-141.

[5] F. Jensen, *An Introduction to Bayesian Networks*. Springer, 1996.

[6] D. Nyang, and J. Song, "Knowledge-proof Based Versatile Smart Card Verification Protocol", *Proc. of ACM SIGCOMM*, Stockholm, Sweden, 2000.

[7] L. Li, J. Han, L. Hu, J. Huai, Y. Liu, and L. Ni, "Pseudo Trust: Zero-Knowledge Based Authentication in Anonymous Peer-to-Peer Protocols", *Proc. of 21th*

*International Parallel and Distributed Processing Symposium*, California, USA, 2007.
[8] A. Abdul-Rahman, and S. Hailes, "Supporting Trust in Virtual Communities", *Proc. of International Conference on System Sciences*, Hawaii, USA, 2000.
[9] Y. Wang, and J. Vassileva, "Bayesian Network Based Trust Model", *Proc. of International Conference on Web Intelligent*, Halifax, Canada, 2003.
[10] S. Milgram, "The Small World Problem", *Psychology Today*, vol. 2, pp. 60-67, 1967.
[11] A. Selcuk, E. Uzun, and M. Pariente, "A Repuation-based Trust Management System for P2P Networks", *International Journal of Network Security*, vol. 6, pp. 235-245, 2008.
[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*, San Diego, USA, 2001.
[13] X509, URL: http://www.ietf.org/rfc/rfc2459.txt.
[14] Y. Xia, Y. Zheng, Y. Li, "A Public Grid Computing Framework Based on a Hierarchical Combination of Middleware", *Proc. of International Workshop on Web-Based Internet Computing for Science and Engineering*, Harbin, China, 2006.
[15] T. Beth, "Efficient Zero-knowledge Identification Scheme for Smart Cards", *Proc. of In Advances in Cryptology: EuroCrypt*, Davos, Switzerland, 1988.

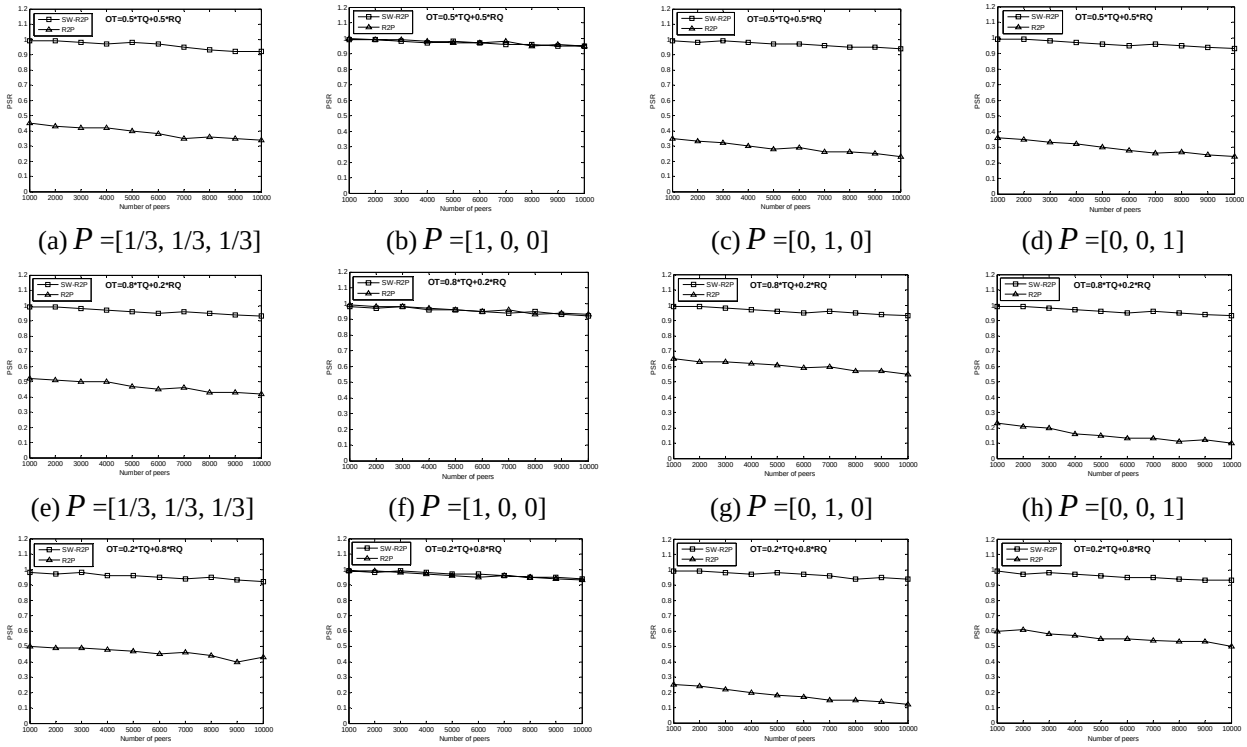
**Yingjie Xia** was born in Fenghua, Zhejiang Province, P.R. China on November 7th, 1982. Now he is a Ph.D. candidate student in the College of Computer Science, Zhejiang University, Hangzhou, Zhejiang, P.R. China. He received his M. Sc. and B. Sc. Degree in the College of Computer Science, Zhejiang University in 2007 and 2004, respectively. His major is computer science and technology.

He deployed the Zhejiang University Campus Grid in 2006 by hierarchically combining the Globus Toolkit and Sun Grid Engine. This work is summarized and published in the proceeding of APWeb, Harbin, 2006. And he also implemented a R2P system, P2P system with role and reputation based access control. The relative paper about this system is published in the proceeding of First International Workshop of Knowledge Discovery and Data Mining, Adelaide, Australia, 2008. Now his work focuses on to improve the R2P system and propose a SW-R2P model, a trusted small world P2P network in Zhejiang University. His research interests cover distributed computing, grid computing and P2P network.

Mr. Xia has got a public patent of his R2P system. He will apply for the patent of the SW-R2P system in China.

**Guanghua Song** is an associate professor of Computer Science and Aeronautics and Astronautics in Zhejiang University, Hangzhou, Zhejiang, P.R. China. He received the Ph.D. degree from the Zhejiang University in 1999. He specializes in operating system, distributed computing and grid computing.

**Yao Zheng** is a Cheung Kong chair professor with Zhejiang University, Hangzhou, Zhejiang, P.R. China, and is directing the Center for Engineering and Scientific Computation, Zhejiang University. He is also serving a deputy dean of the School of Aeronautics and Astronautics, Zhejiang University. He received his Ph.D. degree from the University of Wales Swansea, U.K. He is a senior member of IEEE. His research interests include high performance computing, mesh generation, grid computing and computational visualization.




(a) $P$ =[1/3, 1/3, 1/3] (b) $P$ =[1, 0, 0] (c) $P$ =[0, 1, 0] (d) $P$ =[0, 0, 1]

(e) $P$ =[1/3, 1/3, 1/3] (f) $P$ =[1, 0, 0] (g) $P$ =[0, 1, 0] (h) $P$ =[0, 0, 1]

(i) $P$ =[1/3, 1/3, 1/3] (j) $P$ =[1, 0, 0] (k) $P$ =[0, 1, 0] (l) $P$ =[0, 0, 1]

Figure 7. Peer satisfaction rates (PSR) of the SW-R2P and R2P systems