doi:10.1093/comjnl/bxh000

On Distributed Rating Systems for Peer-to-Peer Networks

Ye Tian*, Di Wu, Kam-Wing Ng

Department of Computer Science and Engineering
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.
** Corresponding author*
Email: {ytian, dwu, kwng}@cse.cuhk.edu.hk

In recent years, many distributed rating systems have been proposed against the increasing misbehaviors of peers in P2P networks. However, the low accuracy, long response time and vulnerabilities under adversary attacks of P2P rating systems have long been criticized and hindering the practical deployment of such a mechanism. There is also a lack of systematic analysis and evaluation for understanding the systems. In this paper, we first present a framework of stochastic analytical model for evaluating P2P rating systems. The performances of two representative designs, namely the unstructured self-managing rating (UMR) system and the structured supervising rating (SSR) system, are then studied with our model. We identify the positive features as well as the negative ones of the two designs with different design choices and under various network environments and adversary attacks. We also propose a configurable loosely supervising rating (LSR) system, and show that this system works inexpensively, and could make tradeoffs between the false rating attack resistance of the UMR system and the accuracy, responsiveness, whitewashing attack resistance as well as failure resilience of the SSR system, thus providing a better overall performance according to the application context.

Received ; revised

1. INTRODUCTION

Peer-to-peer (P2P) networks, which rely on the resources contributed by ordinary users to provide services, have become very popular in recent years. In general, a P2P network is composed of a large number of autonomous peers, and each autonomous peer could choose to behave deceptively instead of being altruistic. For example, in a P2P file-sharing network a peer may upload a corrupted version of the file being requested instead of a good one. For most P2P networks, due to the absence of a centralized management, it is extremely difficult to prevent a peer from behaving deceptively, or to punish the misbehaving peer afterwards. In this paper, we refer to a deceptive peer which aims to subvert the network as **malicious**; and for the rest of the peers, we assume them to be honest and altruistic, and call them the **loyal** peers.

Recently, eBay-like rating systems [1], which work with the basic rationale of sharing a user's past experience, are proposed to be applied on P2P networks for detecting the malicious peers. Generally, for a well designed P2P rating system, it must work inexpensively with accuracy and responsiveness, and should be resilient under network dynamics and be robust under a number of adversary attacks from malicious peers. Unfortunately, due to the decentralized nature of P2P networks, these requirements are not as easily satisfied as on the online marketplaces such as eBay. More importantly, there is a lack of systematic studies, especially theoretical ones, on evaluating how well a P2P rating system could achieve these design objectives.

In this paper, we have developed a framework of stochastic model for understanding P2P rating systems, and have studied the two representative system designs, namely the unstructured self-managing rating (UMR) system and the structured supervising rating (SSR) system with it. We have defined two important metrics for the system performance: the false positive probability and the convergence time. The former represents the accuracy of a rating system while the latter measures the system's responsiveness. We demonstrate how various design choices, network environments and adversary attacks could influence the

performance of the two designs differently. Concretely, our findings include:

- Under the ideal condition without network dynamics and any adversary attacks, the SSR system is more accurate and more responsive than the UMR system, and it scales much better with large-sized networks than the UMR system;
- The overall performance of both P2P rating systems could get improved persistently by increasing peers' cooperation level in sharing their ratings, thus an incentive mechanism is essential in enhancing a P2P rating system;
- Whitewashing attacks could degrade the performance of both P2P rating systems remarkably, especially when an attacking peer could whitewash itself frequently; and the SSR system shows a better whitewashing attack resistance than the UMR system in terms of the long term accuracy;
- Network dynamics could influence both P2P rating systems considerably, and the SSR system has a better failure resilience than its UMR counterpart, regarding the long term accuracy;
- Deploying a personalized credibility sub-system could prevent a P2P rating system from being subverted by false rating attacks to a certain extent, but at the cost of worse accuracy and responsiveness; and when attacking peers are powerful and strategic enough, the P2P rating system could still be subverted; moreover, the UMR rating system has a better resistance against intensive false rating attacks than the SSR rating system, in terms of both the accuracy and the responsiveness.

We also propose a configurable loosely supervising rating (LSR) system for P2P networks in this paper. The LSR system could be viewed as a generalization of a rating system in which both the UMR and the SSR designs are its special cases. We find that the LSR system scales well with large-sized networks. More importantly, it provides means to make tradeoffs between the features of the UMR and the SSR rating systems, and could be deployed to yield an optimal overall performance by pre-setting a protocol parameter properly according to network environments, adversary attacks and application requirements.

We organize the remainder of this paper as follows. In Section 2, we introduce the design space of P2P rating systems, and discuss the factors which are critical to a P2P rating system's performance; in Section 3, a framework of stochastic model is proposed for understanding P2P rating systems, and we study the two representative designs, namely the UMR and the SSR systems with the model; we present and discuss our numerical evaluation results based on the analytical studies in Section 4, and verify the model with simulation experiments; in Section 5, the LSR rating system is proposed and numerically evaluated; Section 6 discusses the related works and finally we conclude the paper in Section 7.

2. BACKGROUND

2.1. P2P Rating Systems: Design Space

2.1.1. Rating Modeling and Aggregation

For most P2P applications, typically two peers are involved in an interaction: the peer which provides the service and the peer which requests and consumes the service. In a P2P rating system, it is required that after each interaction, the service consuming peer should issue a rating on the behavior of the service provider. A peer will use its own historical ratings and the ratings from other peers to evaluate the trustworthiness of its future potential service providers.

One basic issue for building a quantitative rating system is how to model the interaction results into ratings and how to aggregate the ratings to form a trust belief. There are many design options. For rating modeling, ratings could simply be in discrete ranks, or the service consuming peer may use continuous values for expressing the quality of service or the degree of satisfaction. For rating aggregation, the ratings may simply be summed up or averaged to yield a trust score, or advanced probabilistic or possibilistic approaches may be applied. However, the rating modeling and aggregation component is usually independent of other components in a P2P rating system.

2.1.2. Rating Filtering

Without a centralized management, P2P rating systems are vulnerable when false ratings are deliberately issued by malicious peers, thus a mechanism for filtering out the false ratings is necessary. Most of the rating filtering mechanisms are based on a personalized credibility sub-system, in which a peer keeps credibility records on other peers and associate the credibility of a rating with its issuing peer's credibility. We refer to the credibility sub-system "personalized", as for the same target peer, its credibility in different peers' credibility sub-systems may be different.

In a typical credibility sub-system, a peer must issue lots of accurate ratings which are informative to many other peers in order to be recognized as a credible rating issuer by these peers. However, this is hard to achieve in P2P networks: as for most peers, their interaction histories are very limited, hence they are incapable of giving ratings on many different service providers. Conversely, it is also hard for a peer to know many other peers as credible rating issuers in its personalized credibility sub-system, although there are many honest peers existing in the network. Another related issue is that even with a personalized credibility sub-system, a lying peer still has chances to be regarded as credible, for example, it could actively issue honest ratings on some irrelevant peers to build up its credibility, and lies

on the malicious peers it is supposed to cover in some critical interactions.

2.1.3. Rating Management and Distribution

For deploying a rating system on P2P networks, it is essential to design a mechanism for storing, querying and retrieving the rating information. Currently, there are two representative designs for rating management and distribution: the unstructured self-managing rating (UMR) system and the structured supervising rating (SSR) system.

First we describe the UMR design. In this approach, after each interaction, the service consuming peer issues a rating on the service providing peer, and keeps the rating locally. A peer could keep the ratings it has issued by itself, or the super peer in a two-layered unstructured overlay such as Gnutella [2] or KaZaA [3] could keep all the ratings issued by its leaf peers on their behalves. When another peer is interested in this particular service provider, it queries for the ratings on the service provider by looking up the rating keeping peers, using a flooding or a random walk algorithm on the unstructured P2P overlay. Usually the number of the peers visited in a rating query process is constrained by the query's range (e.g. the flooding message's TTL value), and the associated network overhead will be a major concern in a UMR-like rating system.

For the SSR design, unlike in the UMR approach, the ratings are not kept locally, but are managed by a third party. Typically for a peer in a SSR-like rating system, another peer is randomly designated as its supervising peer: when there is a rating issued on this peer, the rating should be sent to its supervising peer; and when another peer is interested in the service provided by this particular peer, it could contact the peer's supervising peer for all the ratings. For the mapping between a peer and its supervisor, a structured P2P overlay algorithm such as Chord [4] or P-Grid [5] may be adopted. With a structured overlay, it could be guaranteed that a peer needs only limited hops to locate the ratings. Our main focus in this paper is to evaluate and compare the two designs of rating management in P2P rating systems.

2.1.4. Client Strategy

A peer in a P2P network needs the rating service when it has obtained a list of service provider candidates from a service query. Usually, the peer queries for the ratings on all the candidates via a P2P rating system, and then chooses one to provide the service. The strategy which a peer follows in selecting its service provider is called its client strategy.

Client strategies could be divided into two major categories: performance centric and trust centric. For the former, a peer puts more emphasis on performance issues, such as the bandwidths in a P2P file-sharing system, than a candidate's trustworthiness obtained from the P2P rating system; while in the latter, a peer prefers to choose a candidate with good trustworthiness for security concerns. Obviously the choice of a client strategy involves a tradeoff between performance and security. And it has been reported that focusing on the trustworthiness only may lead to efficiency problems in P2P networks [6]. It has also been found that randomizing the choices of the service provider with probabilities will make the system more robust under strategic attacks [7].

2.2. Critical Factors

2.2.1. Whitewashing Attacks

In a P2P rating system, ratings and trust beliefs are associated with peers' identities, so it is possible for a peer to get rid of its bad records by abandoning its current identity and rejoining the system with a brand new one. Such an attack is widely known as a "whitewashing attack". Unfortunately, whitewashing attacks are easy to be launched by malicious peers in most P2P networks: as in these networks, cheap pseudonyms are used as the peer identities. However, we will demonstrate in Section 4.2 that whitewashing attacks could only help malicious peers to avoid having bad reputation, but can not enable them to obtain a good one.

2.2.2. Network Dynamics

In a P2P rating system, the ratings are kept by ordinary peers: In an UMR-like system, it is the former service consuming peers or their super peers that keep the ratings, and in a SSR-like system, the ratings are kept by the service providers' supervising peers. However, P2P networks are noted for their dynamics with peers joining and leaving frequently all the time. Thus the availability and the quality of the rating service are influenced when the peers that keep the rating information leave the system abruptly. However, we will demonstrate in Section 4.3 that network dynamics can only make a P2P rating system less effective, but will not sabotage it.

2.2.3. False Rating Attacks

As a P2P rating system relies on ratings to detect the misbehaving peers, one effective way for malicious peers to attack the rating system is to lie with false ratings. Typically malicious peers could launch two types of false rating attacks: ballot stuffing and bad mouthing. In the former, they issue dishonest good ratings on themselves or their colluding peers, aiming to upgrade their trustworthiness dishonestly; while in the latter, dishonest bad ratings are issued on innocent loyal peers, trying to make the rating system less effective in finding trustworthy peers, and further reduce the user's trust on the usability of the entire rating system. As our main concern for rating systems is to detect the malicious peers, we just focus on the ballot stuffing attack and

simply refer to it as false rating attack for the remaining part of this paper.

Usually two kinds of malicious peers could be involved in a false rating attack: the Sybil peers [8] and the colluding peers. In the attacks with Sybil peers, a malicious user controls several peers by registering with multiple identities. Typically, one peer would be used to perform malicious interactions and the other fake peers would issue and spread favorable ratings on it. In the attacks with colluding peers, more than one malicious peers collaborate to attack the rating system. A typical strategy for them is to issue favorable ratings mutually. Obviously, attackers in false rating attacks could combine these two schemes in practice. In the rest of this paper, we do not differentiate the Sybil peers and the colluding peers, but simply refer to them as malicious colluding peers.

3. SYSTEM MODEL AND ANALYSIS

3.1. Analytical Modeling Framework

In this section, we present an analytical modeling framework, which incorporates all the design components previously discussed, to understand and evaluate P2P rating systems. As our main concern is to detect the malicious peers, we divide the ratings in our modeling framework into two categories: the positive ratings and the negative ratings. The former are the ratings stating that the service provider is non-malicious during the interaction, while the latter ratings state that the service provider behaves maliciously in the concerned interaction. Note that for this categorization, any specific rating modeling technique could be adopted as long as the ratings can tell the malicious behaviors from the non-malicious ones. We use the term "amount" to denote the aggregated result of the ratings.

We begin to develop the modeling framework by considering, after a malicious peer, say P_M, joins in a P2P network with N peers, how a rating system could detect it. In our modeling framework, we assume that the P2P rating system is executed in intervals. For the malicious peer P_M, at interval t, the amounts of the positive and the negative ratings issued on it in the entire system are denoted as $n_{pos}(t)$ and $n_{neg}(t)$ respectively. However, when a peer queries for ratings on P_M, only a fraction of the global ratings could be accessed. We use $n'_{pos}(t)$ ($n'_{pos}(t) \leq n_{pos}(t)$) and $n'_{neg}(t)$ ($n'_{neg}(t) \leq n_{neg}(t)$) to denote the amounts of the positive and the negative ratings accessible to a querying peer at time t, and use D_{pos} and D_{neg} to express the relationships between the global and the accessible ratings, namely $n'_{pos}(t) = D_{pos}(n_{pos}(t))$ and $n'_{neg}(t) = D_{neg}(n_{neg}(t))$. Obviously, D_{pos} and D_{neg} are determined by the P2P rating system's rating management and distribution mechanisms. Finally, note that a peer could only use the ratings accessible to itself instead of the global ones to derive its trust beliefs and make decisions.

In P2P networks, peers' behaviors are dynamic, for example, a loyal peer may start to behave maliciously and vice-versa. Due to these dynamics, an observation made a long time ago should not be of the same importance as a recent one in P2P rating systems. As in many systems (e.g. [9] and [10]), we believe that the amount of ratings should decay with time, and use γ_{pos} and γ_{neg} ($0 < \gamma_{pos}, \gamma_{neg} < 1$) as the decaying factors for the positive and the negative ratings respectively. With the decaying factors, n_{pos} or n_{neg} amount of positive/negative ratings at time t_1 decays as $\gamma_{pos}^{t_2-t_1} n_{pos}$ or $\gamma_{pos}^{t_2-t_1} n_{neg}$ amount of positive/negative ratings at time t_2 ($t_2 > t_1$).

For detecting the malicious peer P_M, an important metric is the false positive probability, which is defined as the probability that P_M is regarded as non-malicious by a peer using the P2P rating system. From this definition, we can see that for P_M, its false positive probability at time t, denoted as $p_{fpos}(t)$, is the probability that it gets trusted, i.e., $p_{fpos}(t) = Tr(n'_{pos}(t), n'_{neg}(t))$, where Tr is the procedure for a peer to derive its trust belief from its accessible ratings $n'_{pos}(t)$ and $n'_{neg}(t)$.

After each interval, if P_M gets trusted and is chosen as a service provider, its behavior will change the amounts of the positive or the negative ratings on it. If P_M behaves non-maliciously in the interaction, a positive rating will be issued by the service consumer; on the other hand, if P_M performs a malicious interaction, a negative rating will be issued consequently. Obviously, given P_M's behaviors, how the amounts of the ratings get changed depends on whether or not P_M is chosen to provide service, which is further determined by the trust belief on P_M held by the service requesting peer (i.e. the false positive probability $p_{fpos}(t)$). We use H_{pos} and H_{neg} to denote the changes of the positive and the negative ratings respectively, namely, for P_M, the positive and negative ratings newly issued at interval t are $H_{pos}(p_{fpos}(t))$ and $H_{neg}(p_{fpos}(t))$ respectively.

Summarizing all the discussions, we have the following modeling framework for the evolvement of a P2P rating system as

$$\begin{cases} n_{pos}(t) = n_{pos}(t-1)\gamma_{pos} + H_{pos}(p_{fpos}(t-1)) \\ n_{neg}(t) = n_{neg}(t-1)\gamma_{neg} + H_{neg}(p_{fpos}(t-1)) \\ n'_{pos}(t) = D_{pos}(n_{pos}(t)) \\ n'_{neg}(t) = D_{neg}(n_{neg}(t)) \\ p_{fpos}(t) = Tr(n'_{pos}(t), n'_{neg}(t)) \end{cases} \tag{1}$$

for $t \geq 1$. Note that the analytical modeling framework is very generic without any strong assumptions. We will develop a more concrete model based on this framework for the P2P rating systems discussed in the next section.

3.2. The Rating System Model

As discussed in Section 2.1, there are many design options for each component of a P2P rating system.

In this paper, we do not intend to develop an encyclopedical performance evaluation on all the design option combinations, but to focus on some key components. In particular, we concentrate on evaluating the rating management and distribution mechanism, and study how the designs of the UMR-like and the SSR-like systems behave under different network environments and adversary attacks. To serve this purpose, in this section we present two simple but representative P2P rating systems which differ only in the rating management and distribution mechanism for analytical studies.

In our P2P rating systems, we assume that an evidential rating scheme is adopted. The evidential rating scheme works with a public key infrastructure (PKI), in which each peer is associated with a public/private key pair, and a peer's public key is globally accessible. Before an interaction, the service providing peer P_A issues an evidential rating body message, signs with its private key sk_A and sends it to the service consuming peer P_B. The rating body message contains the identities of the service provider and the service consumer, as well as the interaction details such as the time and the content, and is in the form of $(P_A, P_B, InteractionDetail)_{sk_A}$. After the interaction, the service consuming peer P_B rates P_A's service, appends the rating to the signed rating body message, and signs it with its own private key sk_B to make a complete evidential rating. The message is in the form of $((P_A, P_B, InteractionDetail)_{sk_A}, InteractionRating)_{sk_B}$. The advantage of evidential rating is that neither party of an interaction nor a third party could forge a rating individually, while the service consumer still has the freedom to rate the interaction.

For modeling interaction results, we choose a binary rating scheme, in which one unit amount of positive/negative ratings is issued as the consequence of a non-malicious/malicious interaction. We use the binary rating scheme for two reasons: first, as our main concern is to detect the malicious peers, it is essential to tell the malicious interactions apart from the non-malicious ones, but how "good" those non-malicious interactions are is not important; second, the binary rating scheme is representative and widely used in many systems (e.g. [9], [11] and [12]). Note that although we model the interaction results as binary, it is not necessary that the trust beliefs held by peers are also binary.

For aggregating the ratings and deriving a trust belief, we use the Beta trust model, which is a probabilistic rating aggregating technique widely applied in binary rating systems (e.g. [9] and [13]). Concretely, the Beta trust model states that the probability of a peer to behave non-maliciously in the next interaction follows a Beta distribution as

$$B(p : n'_{pos}, n'_{neg}) = \frac{\Gamma(n'_{pos} + n'_{neg} + 2)}{\Gamma(n'_{pos} + 1)\Gamma(n'_{neg} + 1)} p^{n'_{pos}} (1-p)^{n'_{neg}}$$

where $\Gamma()$ is the Gamma function, and n'_{pos} and n'_{neg} are the amounts of the accessible positive and negative ratings respectively. It is easy to show that the expectation of p with the Beta distribution is $\mu(n'_{pos}, n'_{neg}) = \frac{n'_{pos}+1}{n'_{pos}+n'_{neg}+2}$, and its standard deviation is $\sigma(n'_{pos}, n'_{neg}) = \frac{1}{n'_{pos}+n'_{neg}+2}\sqrt{\frac{(n'_{pos}+1)(n'_{neg}+1)}{n'_{pos}+n'_{neg}+3}}$. Note that when $n'_{pos} = n'_{neg}$, we have $\mu(n'_{pos}, n'_{neg}) = 0.5$, indicating that the trust model has no idea on the peer's trustworthiness; when $n'_{pos} > n'_{neg}$, we have $\mu(n'_{pos}, n'_{neg}) > 0.5$, which means that the peer is more likely to behave non-maliciously, and the higher the value of $\mu(n'_{pos}, n'_{neg})$ is, the more trustworthy the peer will be; we can have a converse result for the condition of $n'_{pos} < n'_{neg}$.

We assume that a peer derives its trust belief as:

$$Tr(n'_{pos}, n'_{neg}) = \mu(n'_{pos}, n'_{neg}) + \eta\sigma(n'_{pos}, n'_{neg}) \quad (2)$$

where η is a optimistic/pessimistic degree in the range of $[-1, 1]$. When $\eta = 0$, the peer is subjective by using the expectation μ as its trust belief; and when $0 < \eta \leq 1 (-1 \leq \eta < 0)$, the peer is optimistic(pessimistic) by taking the positive(negative) deviation into its trust, and the larger(smaller) the η is, the more optimistic(pessimistic) the peer will be. Note that under the Beta trust model, $\sigma(n'_{pos}, n'_{neg})$ decreases rapidly with the increasing n'_{pos} and n'_{neg}, meaning that the trust model is more confident when having more amount of ratings available as references.

The only difference in the two P2P rating systems lies in their rating management and distribution mechanisms. We consider both the UMR-like and the SSR-like designs: the UMR-like system is deployed on an two-layered unstructured P2P overlay such as Gnutella [2], and each super peer is responsible for keeping and distributing the ratings issued by its leaf peers; while the SSR-like system adopts a structured P2P overlay such as Chord [4] as its underlying overlay for mapping a peer to its supervising peer. For the remaining part of this paper, we simply refer to them as the UMR rating system and the SSR rating system for briefness.

Finally, the service consuming peers in our systems adopt a performance centric client strategy. In this strategy, a peer will sort all the service provider candidates in a descending list according to their service providing capacities, and chooses one candidate with the probability that it gets trusted, from the beginning of the list. For example, suppose that the candidate list contains peers as $\{P_1, P_2, ...\}$, and the service requesting peer holds trust belief on peer P_i as Tr_i, then it will choose P_1 with a probability of Tr_1; it will choose P_2 when P_1 is not selected, at a probability of $(1-Tr_1)Tr_2$; and it will consider P_3 if neither P_1 nor P_2 is selected; and so on. Under this strategy, if the malicious peer P_M declares superior service providing capacities and gets itself ranked at the beginning of the candidate list all the

time, then it will be selected as the service provider with the rating system's false positive probability $p_{fpos}(t)$ at interval t.

3.3. Performance Analysis under the Ideal Condition

We first consider an ideal condition for P2P rating systems. The ideal condition is defined as a situation in which a peer never departs the network, and a malicious peer does not launch any kinds of attacks against the rating system, but just performs malicious interactions each time it is selected as the service provider. Obviously under the ideal condition, for a malicious peer P_M, there will be no positive ratings on it, thus $n_{pos}(t) = n'_{pos}(t) = 0$; and each negative rating issued is the consequence that a peer mistakenly selects P_M to provide service, thus the amount of the newly issued negative ratings at interval t is $H_{neg}(p_{fpos}(t)) = c \cdot p_{fpos}(t) = c \cdot Tr(0, n'_{neg}(t))$, here c is the times that P_M appears in other peers' service query results per interval on average, which is a constant. Without loss of generality, we assume $c = 1$.

For deriving $n'_{neg}(t)$, first we consider the UMR system. When a peer queries for ratings, it contacts its super peer, and the super peer uses a flooding or a random walk algorithm to look for more ratings. We assume that during each rating query, S distinct super peers are visited and each is connected to $K-1$ distinct leaf peers, therefore ratings from $S \cdot K$ out of the total N peers could be accessed. However, it is possible that some peers may not be willing to report their ratings to their super peers, due to some reasons such as privacy. Therefore, we use p_c as the probability that a loyal peer will cooperate by reporting its ratings. If there are $n_{neg}(t)$ global negative ratings on the malicious peer P_M at time t, a rating querying peer is expected to access $n'_{neg}(t) = \frac{SK}{N} p_c n_{neg}(t)$ negative ratings in its rating query on average. On the other hand, for the SSR rating system, as all the ratings on P_M are kept by its supervising peer, the rating querying peer only needs to contact P_M's supervising peer for all the ratings. However, we also assume that not all the loyal peers will report their ratings to the supervising peer, but just cooperate at the probability of p_c, so we have $n'_{neg}(t) = p_c n_{neg}(t)$ for the SSR system.

Combining the above analysis, we could rephrase the modeling framework in Equation (1) for the UMR and the SSR rating systems under the ideal condition as

$$\begin{cases} n_{pos}(t) = n'_{pos}(t) = 0 \\ n_{neg}(t) = n_{neg}(t-1)\gamma_{neg} + p_{fpos}(t-1) \\ n'_{neg}(t) = \begin{cases} \frac{SK}{N} p_c n_{neg}(t), & UMR \\ p_c n_{neg}(t), & SSR \end{cases} \\ p_{fpos}(t) = Tr(0, n'_{neg}(t)) \end{cases} \quad (3)$$

for $t \geq 1$.

The above stochastic model indicates that there are two factors influencing the evolvement of the negative ratings: first, when there are few negative ratings in the system, the false positive probability is high and P_M is likely to be selected as a service provider, which consequently causes more negative ratings to be issued; on the other hand, the negative ratings decay with time. Clearly this correlation will lead to a stable equilibrium. If we assume that peers are subjective in deriving their trust beliefs with $\eta = 0$ in Equation (2), by applying the equilibrium condition of $n_{neg}(t-1) = n_{neg}(t)$, we could obtain the amount of the negative ratings at equilibrium, and further derive the equilibrium false positive probability as

$$\bar{p}_{fpos} = \frac{1}{1 + \sqrt{1 + \frac{SKp_c}{N(1-\gamma_{neg})}}} \quad (4)$$

for the UMR rating system and

$$\bar{p}_{fpos} = \frac{1}{1 + \sqrt{1 + \frac{p_c}{1-\gamma_{neg}}}} \quad (5)$$

for the SSR rating system.

From the close-formed solutions, we find that the equilibrium false positive probability achieved by the SSR system is lower than that of the UMR system, and it is independent of the network size N. The close-formed solutions also indicate that the cooperation probability p_c and the decaying factor γ_{neg} are important to the equilibriums of the rating systems. We will numerically discuss the ideal condition performance in Section 4.1.

3.4. Performance Analysis under Whitewashing Attacks

In this section, we study how whitewashing attacks could influence the UMR and the SSR rating systems. As in the previous analysis, we assume that a malicious peer does not launch any other attacks against the P2P rating systems except for the whitewashing attack, and a loyal peer never departs the network or change its identity.

For a particular malicious peer P_M playing whitewashing attacks, we first consider a probabilistic attack strategy, where P_M chooses to whitewash itself with a new identity after each interval at a probability of p_w. Under this probabilistic attack, if we define the consecutive intervals in which P_M uses the same identity as one "session", then the session length follows a geometric distribution as $\Pr[n \text{ intervals in a session}] = p(n) = p_w(1-p_w)^n$. Since in P2P networks, peers are identified with their identities, each time P_M whitewashes itself, the rating system will start to trace it with the initial conditions of $n_{pos} = 0$ and $n_{neg} = 0$, and evolves following the stochastic process described in Equation (3), until P_M whitewashes itself again the next time.

Obviously, when p_w is large, P_M may change its identities frequently. In such a case, considering the equilibrium of a rating system is meaningless as the system may never converge before P_M whitewashes. Instead of using the equilibrium false positive probability, we consider a time-averaged false positive probability $\tilde{p}_{fpos}$, which is defined as the average of P_M's false positive probabilities during all the intervals, to evaluate the long term performance of the P2P rating system under whitewashing attacks.

For calculating the rating systems' time-averaged false positive probabilities, two stochastic processes are under consideration: one is the process that the UMR or the SSR rating system evolves as described in Equation (3), and the other is the process that P_M whitewashes itself, as described by $p(n)$. Palm calculus [14] is used to integrate the two processes. Concretely, according to the Inversion formula [14], the time-averaged false positive probability could be expressed as $\tilde{p}_{fpos} = \frac{E^0[\sum_{t=0}^{n_1-1} p_{fpos}(t)]}{E^0[n_1]}$, where $E^0[\cdot]$ is the Palm expectation and n_1 is the length of the first session. Note that $E^0[\sum_{t=0}^{n_1-1} p_{fpos}(t)] = \sum_{n_1=1}^{\infty} p(n_1) \sum_{t=0}^{n_1-1} p_{fpos}(t)$ and $E^0[n_1] = \sum_{n_1=1}^{\infty} n_1 p(n_1) = \frac{1-p_w}{p_w}$. Combining them, the UMR or the SSR rating system's time-averaged false positive probability under the probabilistic whitewashing attack with attacking probability p_w could be expressed as

$$\tilde{p}_{fpos} = \frac{p_w}{1-p_w} \sum_{n_1=1}^{\infty} p(n_1) \sum_{t=0}^{n_1-1} p_{fpos}(t) \qquad (6)$$

where $p_{fpos}(t)$ evolves following the process described in Equation (3) for the UMR or the SSR rating system.

Another whitewashing attack strategy for the malicious peer P_M is that it does not attack with probabilities, but whitewashes itself after it has conducted a certain number of malicious interactions. Suppose P_M changes its identity after conducting m malicious interactions, then each session of P_M is of a fixed length t_m satisfying $\sum_{t=0}^{t_m} p_{fpos}(t) = m$, and the time-averaged false positive probability could be expressed as

$$\tilde{p}_{fpos} = \frac{1}{1+t_m} \sum_{t=0}^{t_m} p_{fpos}(t) = \frac{m}{1+t_m} \qquad (7)$$

Here $p_{fpos}(t)$ evolves following the process described in Equation (3) for the UMR or the SSR rating system.

We will numerically evaluate the whitewashing attack resistance for different P2P rating systems in Section 4.2.

3.5. Performance Analysis under Network Dynamics

In our previous analysis, we assume that a peer will never depart a P2P network. However, P2P networks are noted for their dynamics, where peers join and leave frequently all the time. In this section, we study how the UMR and the SSR rating systems get influenced by the dynamics of P2P networks. In the following analysis, we assume that a loyal peer departs the network following a lifetime model, while a malicious peer never departs, and there is no adversary attacks against the rating systems.

In general two models are used to describe peers' lifetimes in P2P networks: the exponential lifetime model and the Pareto lifetime model. The exponential lifetime model states that the CDF of a peer's lifetime l follows an exponential distribution as $\Pr(l < x) = 1 - e^{-\lambda x}$. This lifetime model is usually used to describe the behaviors of memoryless peers, where how long a peer is going to stay in the system is independent of the time it has already stayed. Under the exponential lifetime model, we could have the peer lifetime expectation as $E[l] = \frac{1}{\lambda}$ and the peers' departure rate as $\theta = \lambda$ according to Little's law [15]. On the other hand, the Pareto lifetime model states that the CDF of a peer's lifetime l follows a shifted Pareto distribution as $\Pr(l < x) = 1 - (1 + \frac{x}{\beta})^{-\alpha}$ [16]. Under this lifetime model, the longer a peer stays, the less likely it is going to depart. Pareto lifetime is widely observed in P2P networks in recent years [17], and is very suitable to describe a P2P network with a rating system deployed, where the longer a loyal and contributing peer stays, the more reluctantly it is going to depart for the good reputation it has accumulated. Similarly, we could have $E[l] = \frac{\beta}{\alpha-1}$ and $\theta = \frac{\alpha-1}{\beta}$ under the Pareto lifetime model. Both models will be considered in our analysis.

We first study the UMR rating system. In such a system, a rating is kept and distributed by the super peer of its issuer. Assume that in the UMR system, a peer replicates all the ratings it has issued, and when its super peer departs, it reports them to the new super peer designated by the overlay protocol. With such a rating recovery mechanism, a rating is lost only when both the super peer keeping it and the peer issuing it have left. As a portion θ of the peers depart the network each interval, and if we assume that the selection of a super peer is not related to a peer's likelihood of departing, then after an interval, a portion θ^2 of the ratings will be lost on average. Based on the ideal condition model in Equation (3), we could have a revised performance model for the UMR rating system under network dynamics as

$$\begin{cases} n_{pos}(t) = n'_{pos}(t) = 0 \\ n_{neg}(t) = n_{neg}(t-1)(1-\theta^2)\gamma_{neg} + p_{fpos}(t-1) \\ n'_{neg}(t) = \frac{SK}{N} p_c n_{neg}(t) \\ p_{fpos}(t) = Tr(0, n'_{neg}(t)) \end{cases} \qquad (8)$$

for $t \geq 1$, and if we assume that peers derive their trust beliefs subjectively, then the UMR system's equilibrium false positive probability could be obtained by solving

the model as

$$\bar{p}_{fpos} = \frac{1}{1+\sqrt{1+\frac{SKp_c}{N(1-(1-\theta^2)\gamma_{neg})}}} \quad (9)$$

We then consider the SSR rating system. For a particular malicious peer P_M, when its supervising peer departs, all the ratings issued on it will be lost, and the P2P rating service on P_M will be unavailable. When a new supervising peer is assigned for P_M, the rating system on it will restart with the initial condition of $n_{pos} = 0$ and $n_{neg} = 0$. Similar to the discussion on whitewashing attacks, it is meaningless to consider the equilibrium of a rating system under the frequent failures of supervising peers as they may depart before the system converges, therefore we consider the time-averaged false positive probability instead of the equilibrium one to evaluate the SSR rating system under network dynamics. The continuous time Palm calculus [14] is used here to integrate the stochastic processes of the SSR rating system's evolvement and the supervising peers' departure/replacement. We assume that when a peer's supervising peer departs, another peer is designated as its new supervising peer immediately. This assumption is reasonable, as compared with the interaction intervals in P2P networks, the re-stabilize operation of a structured overlay usually takes much less time. Note that under the SSR system, a peer is assigned as a supervising peer some time after it has joined the network, so the time it serves as a supervising peer is actually its residue lifetime between the supervisory designation and its departure. According to [16], we could have the CDFs of the residue lifetime l_r as $\Pr(l_r < x) = 1-e^{-\lambda x}$ and as $\Pr(l_r < x) = 1-(1+\frac{x}{\beta})^{1-\alpha}$ for the exponential lifetime model and the Pareto lifetime model respectively.

By applying the continuous time Inversion Formula [14], the time-averaged false positive probability could be calculated as $\tilde{p}_{fpos} = \frac{E^0[\int_0^{T_1} P_{fpos}(t)dt]}{E^0[T_1]}$. Note that $E^0[\int_0^{T_1} p_{fpos}(t)dt] = \int_0^\infty f(x)\int_0^x p_{fpos}(t)dtdx$ and $E^0[T_1] = \int_0^\infty f(x)dx$, where $f(x)$ is the PMF of a peer's residue lifetime according to the lifetime model. Summarizing all these discussions, we could express the time-averaged false positive probability $\tilde{p}_{fpos}$ for the SSR rating system under network dynamics as

$$\tilde{p}_{fpos} = \lambda\int_0^\infty \lambda e^{-\lambda x}\int_0^x p_{fpos}(t)dtdx \quad (10)$$

when peers' lifetimes follow the exponential lifetime model; and $\tilde{p}_{fpos}$ could be expressed as

$$\tilde{p}_{fpos} = \frac{\alpha-2}{\beta}\int_0^\infty \frac{\alpha-1}{\beta}(1+\frac{x}{\beta})^{-\alpha}\int_0^x p_{fpos}(t)dtdx \quad (11)$$

when peers' lifetimes follow the Pareto lifetime model. Here $p_{fpos}(t)$ evolves as in the stochastic process described in Equation (3) for the SSR rating system. We will numerically evaluate the failure resilience in Section 4.3.

3.6. Performance Analysis under False Rating Attacks

In P2P rating systems, as in any other online reputation systems, it is possible for malicious peers to attack by lying with false ratings. For filtering out false ratings, many P2P rating systems work with a personalized credibility sub-system. Typically, in such a system, a peer tries to keep a personalized credible peer set, and considers the peers in it as credible rating issuers by trusting the ratings from them; for the ratings issued by peers outside of the credible peer set, they are regarded as unreliable. In this section, we consider a false rating attack in which malicious peers collaborate to issue false positive ratings on themselves, and study the UMR and the SSR rating systems working with a personalized credibility sub-system under this attack. As in our previous analysis, we assume that there is no other attacks against the rating systems except for the false rating attack, and a peer never departs the network.

For a particular malicious peer P_M, suppose it has M malicious colluding partners collaborating to issue false positive ratings on it. We assume that for each malicious colluding peer, on average it could manage to keep an amount f of false positive ratings on P_M in the rating system, therefore $n_{pos}(t) = f \cdot M$. We believe f is limited as the ratings decay with time, and the super peers or the supervising peers keeping the ratings are not likely to accept ratings issued by the same peer too frequently.

For a peer P_K in the rating system, suppose the average size of its credible peer set is R. A loyal peer will get admitted into P_K's personalized credible peer set by issuing honest ratings actively for P_K, however, a malicious colluding peer may also have its chance to be accepted by actively issuing honest ratings for P_K on some other irrelevant peers. We assume that a malicious colluding peer of P_M will have a relative probability of p_l to be accepted into P_K's personalized credible peer set, compared with the chance of a loyal peer. As P_K's credible peer set contains R peers and there are M malicious colluding peers in the network, on average, P_K will have $R\frac{p_l M}{N-(1-p_l)M}$ malicious colluding peers and $R\frac{N-M}{N-(1-p_l)M}$ loyal peers in its credible peer set.

For a personalized credibility sub-system, how to handle the ratings issued by peers outside of the credible peer set is a problem. We assume that a small weight w ($0 \le w < 1$) will be assigned for those ratings, for example, n amounts of ratings issued by peers outside of the credible peer set will be regarded as equal to $n \cdot w$ amounts of ratings issued by credible peers. As ratings issued by peers in and outside of the credible peer set are assigned with different weights, for the peer P_K querying for ratings on P_M, only a

portion $\xi_1 = \frac{Rp_l}{N-(1-p_l)M} + w(1 - \frac{Rp_l}{N-(1-p_l)M})$ of the false positive ratings and a portion $\xi_2 = \frac{R}{N-(1-p_l)M} + w(1 - \frac{R}{N-(1-p_l)M})$ of the negative ratings it retrieves from the rating system will be considered. Based on the above discussions, we could extend the analytical model presented in Equation (3) for modeling the UMR and the SSR rating systems working with a personalized credibility sub-system under false rating attacks as

$$\begin{cases} n_{pos}(t) = f \cdot M \\ n_{neg}(t) = n_{neg}(t-1)\gamma_{neg} + p_{fpos}(t-1) \\ n'_{pos}(t) = \begin{cases} \xi_1 \frac{SK}{N} n_{pos}(t), & UMR \\ \xi_1 n_{pos}(t), & SSR \end{cases} \\ n'_{neg}(t) = \begin{cases} \xi_2 \frac{SK}{N} p_c n_{neg}(t), & UMR \\ \xi_2 p_c n_{neg}(t), & SSR \end{cases} \\ p_{fpos}(t) = Tr(n'_{pos}(t), n'_{neg}(t)) \end{cases} \quad (12)$$

for $t \geq 1$. We use the denotation $n'_{pos}(t)$ and $n'_{neg}(t)$ here for the amounts of the positive/negative ratings actually taken into consideration when a peer derives its trust belief, and assume that unlike loyal peers, which report their ratings with probability p_c, malicious colluding peers trying to mislead the rating systems will always report their false ratings when queried.

Assuming that peers derive their trust beliefs subjectively, we could obtain the equilibrium false positive probability under false rating attacks as

$$\bar{p}_{fpos} = \frac{1 + \frac{\xi_1 fMSK}{N}}{1 + \frac{\xi_1 fMSK}{2N} + \sqrt{(1 + \frac{\xi_1 fMSK}{2N})^2 + \frac{\xi_2 \frac{SK}{N} p_c (1 + \frac{\xi_1 fMSK}{N})}{1-\gamma_{neg}}}} \quad (13)$$

for the UMR rating system and as

$$\bar{p}_{fpos} = \frac{1 + \xi_1 fM}{1 + \frac{\xi_1 fM}{2} + \sqrt{(1 + \frac{\xi_1 fM}{2})^2 + \frac{\xi_2 p_c (1 + \xi_1 fM)}{1-\gamma_{neg}}}} \quad (14)$$

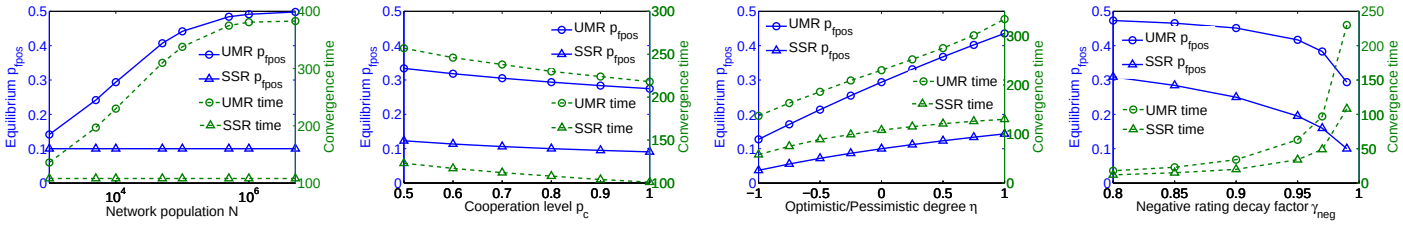
for the SSR system. Note that if we let $w = 0$ and $p_l = 0$, which means that peers do not consider ratings from outside of their credible peer sets and no malicious colluding peer is mistakenly admitted into a peer's credible peer set, then $\xi_1 = 0$ and $\xi_2 = \frac{R}{N}$. With this perfect personalized credibility sub-system, we have $\bar{p}_{fpos} = \frac{1}{1+\sqrt{1+\frac{Rp_c}{N(1-\gamma_{neg})}}}$ for the SSR system. Comparing it with the UMR system's $\bar{p}_{fpos}$ under the ideal condition in Equation (4), we can see that they are identical if we exchange $S \cdot K$ with R. In other words, even with a perfect credibility sub-system, the performance of the SSR system under false rating attacks is degraded as in the UMR system under the ideal condition. We predict that a personalized credibility sub-system can only prevent P2P rating systems from being misled to some extent, but at the cost of worse performance compared with the one under the ideal condition. We will verify this point and numerically evaluate the false rating attack resistance in Section 4.4.

In the above we have developed a stochastic model for the UMR and the SSR rating systems, and discussed the techniques for evaluating the systems' performances with different critical factors one at a time. Our methodology could easily be extended to handle more complex situations when more than one critical factors are present. For example, for the SSR system's performance under both network dynamics and false rating attacks, we could apply Equations (10) and (11) according to peers' lifetime models, and integrate the stochastic process described in Equation (12) for the evolvement of the SSR system under false rating attacks to obtain its time-averaged false positive probability.

4. EVALUATION

The analysis in Section 3 provides a theoretical foundation for evaluation and comparison of the UMR and the SSR rating systems' performances. In this section, we numerically study the two systems under different adversary attacks and network environments. We also validate our analytical model by simulation.

In the UMR rating system, a two-layered unstructured overlay is adopted for implementing the rating system, while for the SSR system, a Chord-like structured overlay is used. As the two designs are diverse in many aspects, in order to make comparisons, we need to conduct the performance evaluation based on some fair principle. A reasonable choice is that both systems work with the same network overhead. Before we study and compare the performances of the rating systems in detecting the malicious peers, it is necessary to understand the network overhead for each design. For the UMR system deployed on a two-layered unstructured P2P overlay, only the super peers are involved in the rating querying and transmission procedures, and the super peer overlay could be viewed as a flat unstructured overlay. Assuming that a super peer floods the query on the super peer overlay for ratings, then for visiting S distinct super peers, a network overhead of $O(S)$ is required; on the other hand, for the SSR system, if we assume Chord [4] is used here, then for reporting and retrieving a peer's ratings, a network overhead of $O(\log_2 N)$ is required. For fairness, we assume that in the UMR system, the rating querying peer's super peer visits $S = \min(\frac{N}{K}, \log_2 N)$ super peers during the flooding, at a network overhead of $O(\log_2 N)$ when $\log N \leq \frac{N}{K}$. Moreover, we assume that for the UMR system's two-layered overlay, each super peer is connected to 45 distinct leaf peers, as observed in Gnutella [18]. Based on these assumptions, a peer in the UMR system will obtain ratings issued by $S \cdot K = \min(N, 45 \log N)$ distinct peers in each query in our evaluation.



(a) (b) (c) (d)

FIGURE 1. Performances of UMR and SSR with (a) varying network size N; (b) varying cooperation level p_c; (c) varying optimistic/pessimistic degree η; and (d) varying decaying factor γ_{neg}

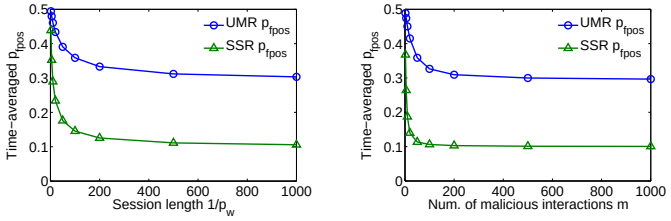
4.1. Ideal Condition Performance

We first study the performances of the UMR and the SSR rating systems under the ideal condition, under which both systems could achieve stable equilibriums, as indicated in Section 3.3. We are interested in two properties of their equilibriums: the equilibrium false positive probability and the convergence time. The former is the metrics for the accuracy of a rating system, while the latter measures the system's responsiveness. Specifically, for the convergence time, we measure the number of the intervals required for a rating system to reach the 98 percentiles close to its equilibrium false positive probability. Previous analysis in Section 3.3 shows that even under the ideal condition, the rating systems do not work perfectly, but are influenced by factors including: 1) network size; 2) peers' cooperation level in sharing their ratings; 3) peers' attitude in forming their trust beliefs; and 4) rating's decaying rate. By applying the analytical results in Equations (3), (4) and (5), we numerically study the influences of these factors and present the results in Figure 1. For the remaining part of this paper, unless otherwise specified, we always set the network size N as $10,000$, peers' cooperation level p_c is set as 0.8, negative ratings decay with a factor γ_{neg} of 0.99, and peers are subjective in deriving their trust beliefs by choosing $\eta = 0$. For each figure, we plot the equilibrium false positive probabilities of the two rating systems as well as their convergence times. Note that there are two y-axes in each figure.

We plot the systems' performances under varying network size N from one thousand to five millions in Figure 1(a). The numerical results show that by retrieving ratings globally, the SSR system has a better performance than the UMR design in both the accuracy and the responsiveness. Moreover, we find that the SSR system scales perfectly, with its equilibrium false positive probability independent of the network size, as indicated by Equation (5); while the UMR system can not scale very well with large sized P2P networks, and is nearly useless by having an equilibrium false positive probability close to 0.5 when the network size is in the millions.

In either the UMR or the SSR design, the rating system relies on the willingness of peers to contribute their ratings. In Figure 1(b), we study the influence of peers' cooperation levels on the performance of the rating systems by varying the cooperation probability p_c from 0.5 to 1.0. From the figure, we find that when peers are more willing to report their ratings, lower equilibrium false positive probabilities and shorter convergence times could be achieved by the rating systems. Our observation indicates that for both rating system designs, providing incentives for peers to contribute their ratings is essential in improving the systems' performances.

In our rating systems, a peer may be subjective, optimistic or pessimistic when deriving its trust belief. In this experiment, we study the performances of the rating systems under varying optimistic/pessimistic degrees η from -1 to $+1$, and plot the results in Figure 1(c). As we can see from the figure, if peers behave pessimistically by setting η negative in Equation (2), lower equilibrium false positive probabilities and shorter convergence times for both the UMR and the SSR systems could be achieved. However, we must point out that the relatively better performance is at the cost of making the system more vulnerable under the "bad mouthing" attack, in which an attacking peer may exploit peers' pessimistic altitude towards negative ratings by issuing false negative ratings on good service providers. When the "bad mouthing" attack is prevalent, peers may have difficulty in finding good service providers, and consequently lose their trust on the entire rating system. We believe that a pessimistic altitude is only necessary when the interactions on a P2P network are critical, in which a peer may prefer having no interactions rather than risking of having malicious ones. As we focus on detecting the malicious peers in this paper, we do not discuss the tradeoff between preserving the user trust of a rating system and improving its performance in detecting the malicious peers in details, but simply assume that peers are subjective by setting $\eta = 0$ in our further evaluation.



(a) (b)

FIGURE 2. Performances of UMR and SSR with (a) varying session length $\frac{1}{p_w}$; and (b) varying num. of malicious interactions m

As we have discussed, due to peers' behavior dynamics, ratings in a P2P rating system should decay with time. For our last study under the ideal condition, we investigate how the decaying factors of ratings influence the UMR and the SSR rating systems' performances. As there is no positive ratings under the ideal condition, we only vary the negative ratings' decaying factor γ_{neg} from 0.8 to 0.999, and plot the results in Figure 1(d). We find that with a larger decaying factor, which means that peers are less forgetful by memorizing ratings longer, both rating systems are more accurate with lower false positive probabilities but less responsive with longer convergence times. We explain this with the fact that when negative ratings decay slowly, it is hard for a malicious peer to be forgotten by the rating system as its behaviors are memorized longer, so it needs to wait longer to have the chance to behave maliciously again. For the extreme case, when the negative ratings never decay, the rating system will accumulate more and more negative ratings and with lower and lower false positive probabilities approaching zero, but will never converge, as indicated by the tendencies shown in the figure when γ_{neg} is close to one. We must point out that a large decaying factor is not appropriate for a P2P rating system, as it makes the system incapable of handling peers' behavior dynamics; moreover, a malicious peer could also exploit peers' unforgetfulness to attack the rating system by "bad mouthing", in which a false negative rating could influence the reputation of the victim peer for a long time. For the remaining part of this paper, we choose $\gamma_{neg} = 0.99$, which we believe is a proper value for a P2P rating system to decay its ratings.

4.2. Performance under Whitewashing Attacks

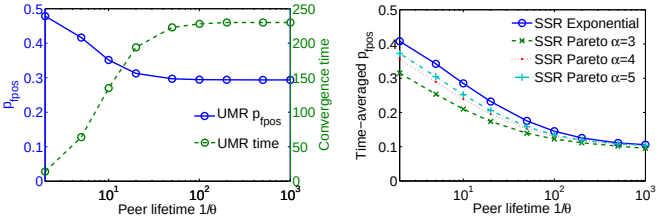
In this section, we investigate the performances of the UMR and the SSR rating systems under whitewashing attacks. Based on the methodology discussed in Section 3.4, we numerically obtain the time-averaged false positive probabilities of the two systems, and plot the results in Figure 2. For the first study, we assume that a malicious peer will whitewash itself with probability p_w after each interval, and use $\frac{1}{p_w}$ as the expected session length for the malicious peer. By applying Equation (6), we plot the time-averaged false positive probabilities under various whitewashing probabilities from 0.001 to 0.5 for both the UMR and the SSR rating systems in Figure 2(a). It is observed that when the session length $\frac{1}{p_w}$ increases, the time-averaged false positive probabilities of both systems approach their ideal condition equilibrium false positive probabilities asymptotically, especially when $\frac{1}{p_w}$ is larger than 200 intervals; on the other hand, when the malicious peer could whitewash itself with high probability p_w, the time-averaged false probabilities increase rapidly, indicating that the intensive whitewashing attacks could influence the P2P rating systems severely.

We also study the attack strategy that a malicious peer whitewashes itself after performing a fixed number m of malicious interactions, and plot the results based on Equation (7) in Figure 2(b), with m varying from 2 to 1,000 interactions. We have observed similar outputs under this attack strategy as under the probabilistic whitewashing attacks, in which the long term performances of the rating systems get degraded seriously when the malicious peer attacks with high frequencies. Note that under either attack strategy, the SSR rating system has a better long term performance than the UMR system. This could be explained with the fact that the SSR system converges towards a lower equilibrium false positive probability with a shorter convergence time than the UMR system in each session, making its time-averaged false positive probability lower than that of the UMR system.

Finally, we conclude that both the UMR and the SSR rating systems are notably influenced by whitewashing attacks, especially when a malicious peer attacks with high frequencies. For safeguarding a P2P rating system against whitewashing attacks, constraining the attacking frequencies of the attacking peers is a feasible approach.

4.3. Performance under Network Dynamics

In this section, we investigate the performances of the UMR and the SSR rating systems under a dynamic network environment. Based on the analysis in Section 3.5, we obtain the numerical results, and present them in Figure 3. We first study the UMR system, in which ratings are kept by super peers. For a particular malicious peer in the UMR system, departures of some peers will only cause a partial loss of the ratings issued on it, and influence the rating system's equilibrium. By applying Equations (8) and (9), we derive and plot the UMR system's equilibrium false positive probabilities and its convergence times in Figure 3(a), when the peer lifetime $\frac{1}{\theta}$ is varied from 2 intervals to 1,000 intervals. Here we do not differentiate the exponential lifetime



(a) (b)

FIGURE 3. (a) Equilibrium false positive probabilities and convergence times for UMR; and (b) time-averaged false positive probabilities for SSR under varying peer lifetime $\frac{1}{\theta}$

model and the Pareto lifetime model, as they have the same effect on the rating system. It is observed from the figure that when the network is highly dynamic with short peer lifetimes, the performance of the UMR system gets degraded seriously, as large amounts of ratings are lost with the departing peers. We also find that as the network becomes more stable with longer peer lifetimes, both the equilibrium false positive probability and the convergence time approach the ideal condition values asymptotically.

We also study the SSR rating system under network dynamics. In such a system, for a particular malicious peer, all the ratings issued on it will be lost with the departure of its supervising peer, making the rating system to re-evolve from the beginning. We apply Equations (10) and (11) to obtain the time-averaged false positive probability for the SSR system under network dynamics, and plot the results in Figure 3(b). In our evaluation, we use the exponential lifetime as well as the Pareto lifetime to model the departures of the supervising peers, and vary the expected lifetime $\frac{1}{\theta}$ from 2 intervals to 1,000 intervals. Again we find from the figure that the SSR rating system performs poorly under a highly dynamic network, but when the network becomes more stable, the time-averaged false positive probabilities under all the lifetime models approach the system's ideal condition performance asymptotically. Another observation is that under the same level of network dynamics, the rating system with the exponential lifetime model has an inferior performance than the systems with the Pareto lifetime model, and the more heavy-tailed the peers' lifetime is (with smaller α), the better the rating system will perform. We explain this as under a more heavy-tailed lifetime model, a few supervising peers will have much longer lifetime than the average, thus contributing greatly to the system-wide time-averaged false positive probability.

Finally, we conclude that it is infeasible to keep either rating system effective under an extremely dynamic network, and if we view the UMR system's equilibrium false positive probability as its time-averaged performance by measuring the system long enough, then by comparing Figure 3(b) with Figure 3(a), we can see that the SSR system has a better performance than the UMR system under the same level of network dynamics. Similar to the discussion on whitewashing attacks, we also believe this owes to the SSR system's better ideal condition performance.

4.4. Performance under False Rating Attacks

In this section, we apply the analytical results in Section 3.6 to study the influence of false rating attacks on the UMR and the SSR rating systems. We assume that a personalized credibility sub-system is available for filtering out false ratings. We consider two different situations regarding the effectiveness of the personalized credibility sub-system. For the first situation, we suppose that a peer's reliable peer set is very large by setting its size as $R = 5,000$ peers in the model of Equation (12). Under this setting, the peer is very confident with its personalized credibility sub-system as it has half of the population as its credible rating issuers. We set the weight assigned for the ratings issued by peers outside of the credible peer set very small as $w = 0.001$ for this situation, as it is not likely for a confident peer to consider the ratings from unreliable peers. We also consider an unconfident situation in which a peer knows very few credible peers by setting its credible peer set size small as $R = 50$ peers. As there are so few ratings available from inside of its credible peer set, an unconfident peer under this condition needs to consider the ratings from outside of its credibility peer set. So we set $w = 0.1$ under this situation.

For each malicious peer, we assume it has a fixed number of colluding peers in the system by setting $M = 100$ peers. We vary their attacking abilities and strategies to study the rating systems under the confident and the unconfident cases. The results are plotted in Figure 4. In Figure 4(a) and (b), we study the confident case. We first vary p_l, the relative probability of a malicious colluding peer to get trusted, from 0.0 to 1.0, while fixing the false rating amount issued per colluding peer as $f = 1$. The equilibrium false positive probabilities and the convergence times for both the UMR and the SSR systems are plotted in Figure 4(a). We can see that both systems' performances get degraded severely with an increasing p_l, and the equilibrium false positive probabilities are higher than 0.5 when p_l is larger than 0.5, meaning that if the malicious colluding peers have half the chance of the loyal peers to get trusted, the false rating attack could successfully mislead the rating systems. We also find that under this case, the SSR system has a worse performance than the UMR system with high p_l values, and gets influenced by the false rating attack more severely. We also consider another attack strategy, in which the malicious colluding peers issue more false

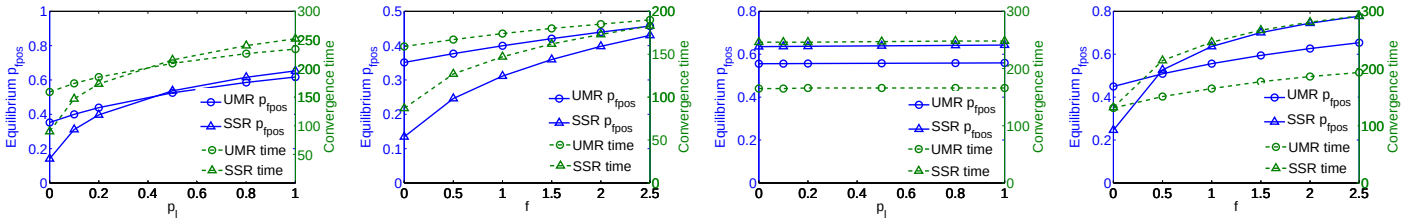


FIGURE 4. Performances of UMR and SSR under (a) varying p_l for confident peers; (b) varying f for confident peers; (c) varying p_l for unconfident peers; and (d) varying f for unconfident peers

ratings by varying f from 0.0 to 2.5, but fixing p_l as 0.1. The results are presented in Figure 4(b). Compared with the previous attack strategy, this attack is relatively less effective, as both rating systems keep their equilibrium false positive probabilities lower than 0.5 all the time, which means that with a confident peer, a powerful and effective credibility sub-system can mitigate the influence of the false rating attacks with the attack strategy of simply issuing more false ratings effectively. And again we find that the SSR system shows inferior attack resistance than the UMR system under this attack.

For the unconfident case, we also study the UMR and the SSR rating systems' performance under the same attack strategies with the same parameter settings. The results are presented in Figure 4(c) and (d). We can see from Figure 4(c) that under the attack strategy of increasing p_l, the equilibrium false positive probabilities of both rating systems are higher than 0.5 all the time, and are almost independent of p_l. The observation indicates that the false rating attacks have misled the rating systems successfully when peers are unconfident, and the main source of the attacks comes from outside of the peer's credible peer set. Note that under this attack strategy, the SSR system has a higher equilibrium false positive probability than the UMR system's all the time. In Figure 4(d), under the attack strategy of increasing f, we can see that both rating systems' performances get degraded with more false ratings released, and the rating systems are misled by the false rating attacks when $f > 0.5$. Again we find that the SSR system shows inferior attack resistance than the UMR system under this attack strategy.

In summary, false rating attacks could degrade the performance of both the UMR and the SSR rating systems to different degrees, depending on the effectiveness of the personalized credibility sub-system and the malicious colluding peers' attacking strategy. Generally speaking, both systems perform worse than under the ideal condition, and malicious colluding peers could attack the rating systems effectively with different strategies according to how powerful and reliable the personalized credibility sub-system is. Another finding

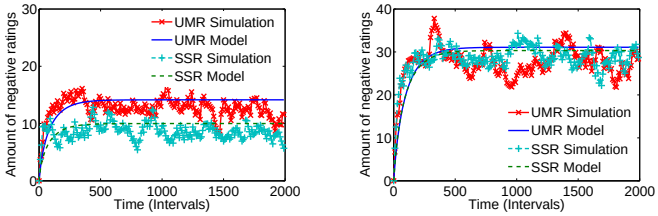


FIGURE 5. Comparison between simulation and modeling results of UMR and SSR with (a) no lying peers, and (b) 10 lying peers

from the evaluation is that the UMR system shows a superior false rating attacks resistance than the SSR system. Finally, we conclude that a P2P rating system can maintain its performance only on condition that its credibility sub-system is powerful enough and reliable enough; otherwise, the system is vulnerable under strategic false rating attacks and could easily be misled.

4.5. Simulation Validation

The results in the previous sections are obtained from numerical studies based on the analytical model in Section 3, thus it is necessary to verify the model's soundness and accuracy. In this section, we simulate a P2P network with 1,000 peers, deploy the UMR and the SSR rating systems upon it separately, and compare the simulation outputs with the modeling results. The metrics under the numerical study is the false positive probability, which can not be observed directly in the simulation. Therefore, in this experiment we measure the amount of the negative ratings accumulated in the system, and compare it with the variable $n_{neg}(t)$ in the model instead. The comparisons are plotted in Figure 5. In Figure 5(a), we simulate a network with only one malicious peer, thus the rating systems are free of false rating attacks. We can see that the modeling results match the simulation outputs very well for both the UMR and the SSR systems. In Figure 5(b), ten malicious colluding peers are inserted into the network,

and we allow a loyal peer to use all the other peers as its credible rating issuers. Again we can see that the modeling results and the simulation outputs are very close, which again verifies the soundness and accuracy of the analytical model.

5. A LOOSELY SUPERVISING RATING SYSTEM

Previous numerical evaluation shows that there are pros and cons for both the UMR and the SSR designs of P2P rating systems, depending on various adversary attacks and network dynamics. Generally speaking, the SSR system is more accurate and responsive than the UMR system, under the ideal condition. When the rating system is under whitewashing attacks or the P2P network is dynamic, the SSR system also outperforms the UMR system for its better ideal condition performance. On the other hand, when malicious peers collude to attack the rating systems with false positive ratings, the SSR system is more likely to be misled than the UMR system, even working with a functional but imperfect personalized credibility sub-system for filtering out the false ratings. Since P2P networks are diverse in terms of interaction sensitivity, network dynamics and adversary power, different properties are desired for the rating systems deployed under different application contexts, thus providing flexibility regarding the accuracy, responsiveness, failure and attack resilience is necessary. In this section, we propose a configurable loosely supervising rating (LSR) system, which could be viewed as a generalization of the UMR and the SSR designs, and show that this system could enable the user to trade off the features of the UMR and the SSR systems by pre-setting a system parameter properly.

5.1. System Model

As in the SSR system, the LSR system works on a structured P2P overlay. In the LSR system, each peer is associated with an identity assigned by the overlay's naming algorithm (e.g. hashing) as $(u_1, u_2, \ldots, u_l)$, here l is the fixed length of the identity string, and u_i is a digit in an alphabet of size Δ. If we first assume that the overlay is fully populated, which means for each identity in the ID space, there is a corresponding peer, then we have $N = \Delta^l$. For a P2P interaction, suppose the identity of the service providing peer is $(p_1, p_2, \ldots p_l)$ and the service consuming peer's identity is $(c_1, c_2, \ldots, c_l)$, then in the LSR system, a rating issued by the service consuming peer will be sent to a peer with the identity of $(c_1, c_2, \ldots, c_i, p_{i+1}, \ldots p_l)$, here i is a system parameter of the protocol. Obviously in this scheme, the ratings on the peer identified as $(p_1, p_2, \ldots p_l)$ will be kept by a set of the peers which has the postfix of $p_{i+1}, \ldots p_l$ in their identities. We call such a set as the jury for the peer of $(p_1, p_2, \ldots p_l)$, and each peer in the jury is called a juror peer. If another peer wishes to access the ratings on the peer of $(p_1, p_2, \ldots p_l)$, it could contact arbitrarily **one** juror peer postfixed with the string $p_{i+1}, \ldots p_l$ in its identity for ratings. Similar to the typical SSR-like protocols (e.g. [19]), we could relax the fully-populated assumption by allowing the rating to be sent to the peer which is responsible for the identity of $(c_1, c_2, \ldots, c_i, p_{i+1}, \ldots p_l)$ in the ID space. We will refer to a LSR rating system with parameter i as an i-LSR rating system in the rest of this paper.

Taking a closer look at the LSR system design, we could find that it may be viewed as a generalization of the UMR and the SSR systems. In the LSR system, if we set $i = 0$, then to which peer a rating should be sent will only depend on its issuing peer, and the UMR system could be viewed as a special case of the 0-LSR system in which the issuing peer just sends the rating to itself. On the other hand, if we set $i = l$, then all the ratings must be kept by the peer they are issued on. And if we apply an extra one-to-one randomized mapping among the peers, and send the ratings to the corresponding peer according to this mapping relationship, it is actually the SSR rating system. Based on this observation, we hypothesize that by adjusting the value of i, the i-LSR rating system could do a tradeoff between the properties of the UMR and the SSR systems. We will verify this point via analysis and numerical evaluation.

5.2. Performance Analysis

Under the ideal condition, for developing the stochastic model on the LSR rating system, we consider a particular malicious peer P_M, whose false positive ratings and negative ratings could be found in its jury with Δ^i juror peers. By contacting one juror peer, a rating querying peer could obtain ratings issued by $\frac{N}{\Delta^i}$ peers on average in the i-LSR rating system. Following the same denotations and assumptions as in Section 3.3, the analytical model for the i-LSR rating system under the ideal condition could be expressed as

$$\begin{cases} n_{pos}(t) = n'_{pos}(t) = 0 \\ n_{neg}(t) = n_{neg}(t-1)\gamma_{neg} + p_{fpos}(t-1) \\ n'_{neg}(t) = \frac{n_{neg}(t)}{\Delta^i} p_c \\ p_{fpos}(t) = Tr(0, n'_{neg}(t)) \end{cases} \tag{15}$$

for $t \geq 1$. Applying the equilibrium condition of $n_{neg}(t) = n_{neg}(t-1)$, we can derive the equilibrium false positive probability for the i-LSR rating system under the ideal condition as

$$\bar{p}_{fpos}(t) = \frac{1}{1 + \sqrt{1 + \frac{p_c}{\Delta^i(1-\gamma_{neg})}}} \tag{16}$$

When the malicious peer P_M attacks the LSR rating system by whitewashing itself, then after each whitewashing, new juror peers will be designated for P_M according to its new identity, and the rating

system needs to re-accumulate the ratings from the beginning. Similar to the discussion in Section 3.4, under whitewashing attacks, we believe that considering the equilibrium of the LSR rating system is meaningless, and instead use the time-averaged false positive probability for evaluating the system. Concretely, we could apply the method based on Palm calculus in Equations (6) and (7) according to the attack strategies, and integrate the stochastic process for the i-LSR system to evolve in Equation (15) to obtain its time-averaged false positive probabilities under different whitewashing attacks.

We also consider the influence of network dynamics on the LSR rating system. In a dynamic network, when a juror peer of the malicious peer P_M departs, the peer assigned as its new juror will re-accumulate the ratings issued from the corresponding ID space region from the beginning. Under the frequent failures of juror peers, it is meaningless to consider the LSR rating system's equilibrium, as a juror peer may depart before the system converges. As in the discussion for the SSR system under network dynamics in Section 3.5, we evaluate the LSR system's performance in terms of the time-averaged false positive probability. If peers are assumed to choose one juror peer randomly each time they query for ratings, we can simply apply the method based on Palm calculus in Equations (10) and (11), to obtain the time-averaged false positive probabilities for the LSR system under different lifetime models, by integrating its evolving stochastic process expressed in Equation (15).

When more than one malicious peers attack the rating system collaboratively by issuing false positive ratings, we again assume that there is a personalized credibility sub-system based on a credible peer set for peers to filter out the false ratings. By following the same assumptions and denotations in Section 3.6, we could develop the analytical model for the i-LSR system under false rating attacks as

$$\begin{cases} n_{pos}(t) = f \cdot M \\ n_{neg}(t) = n_{neg}(t-1)\gamma_{neg} + p_{fpos}(t-1) \\ n'_{pos}(t) = \xi_1 \frac{n_{pos}(t)}{\Delta^i} \\ n'_{neg}(t) = \xi_2 p_c \frac{n_{neg}(t)}{\Delta^i} \\ p_{fpos}(t) = Tr(n'_{pos}(t), n'_{neg}(t)) \end{cases} \tag{17}$$

for $t \geq 1$, here ξ_1 and ξ_2 are of the same meanings as in Section 3.6. Solving the model with the equilibrium condition, we could have the equilibrium false positive probability for the i-LSR rating system under false rating attacks as

$$\bar{p}_{fpos} = \frac{1 + \frac{\xi_1 fM}{\Delta^i}}{1 + \frac{\xi_1 fM}{2\Delta^i} + \sqrt{(1 + \frac{\xi_1 fM}{2\Delta^i})^2 + \frac{\xi_2 \frac{1}{\Delta^i} p_c (1 + \frac{\xi_1 fM}{\Delta^i})}{1-\gamma_{neg}}}} \tag{18}$$

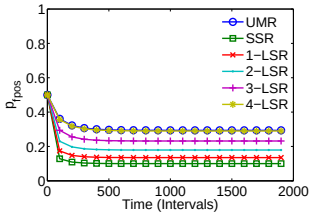


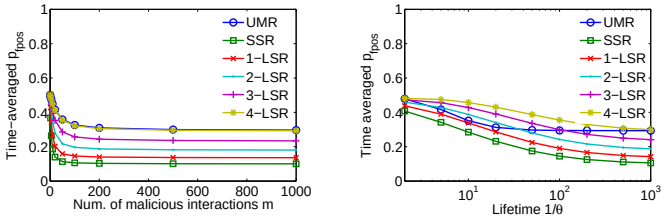
FIGURE 6. The evolution of UMR, SSR and $1, 2, 3, 4$-LSR systems with time

5.3. Performance Evaluation

Based on the theoretical analysis, we numerically evaluate the LSR rating system in this section. As in Section 4, we first discuss the system's network overhead, then we investigate its performance under different network environments and adversary attacks.

For the overhead of the LSR system, as all the ratings on a particular peer are kept by the peers in its jury of size Δ^i, and a rating querying peer contacts only one juror peer in each query, then similar to the SSR system, on average an overhead of $O(\log_\Delta N)$ will be incurred for a peer to query for and retrieve the ratings. If Chord [4] is adopted as the underlying overlay, then we have $\Delta = 2$. Recall that for the i-LSR system, by contacting one juror peer, a querying peer will retrieve ratings issued by $\frac{N}{\Delta^i}$ peers on average, which is equal to the number of the distinct peers visited by a rating query in the UMR system when $S \cdot K = \frac{N}{\Delta^i}$, however, the UMR system works at an overhead of $O(S)$. In other words, the i-LSR system may achieve the approximate performance as the UMR system when $S \cdot K \approx \frac{N}{\Delta^i}$, but works less expensively as long as $O(\log_2 N) < O(S)$. Moreover, the overhead of the LSR system is independent of the system parameter i, meaning that we can adjust the value of i for better overall performance without introducing additional overhead.

For the performance evaluation, we first consider the ideal condition. We use the default parameter settings as in Section 4.1. Figure 6 plots the evolution of the false positive probabilities for different i-LSR systems with $i = 1, 2, 3, 4$, after one malicious peer is inserted. We also plot the curves for the UMR and the SSR systems for comparison. From the figure we can see that the i-LSR systems perform just between the UMR and the SSR systems regarding the false positive probabilities. In detail, we find that with a smaller value of i, the i-LSR system works more similar to the SSR system with lower false positive probabilities and converges faster; while with a larger valued i, the i-LSR system behaves more UMR-like, with higher false positive probabilities and converges slower. Especially, we can see that the 4-LSR system performs very close to the UMR system, as for the 4-LSR system and the UMR system under evaluation, $S \cdot K \approx \frac{N}{\Delta^4}$.



(a) (b)

FIGURE 7. Time-averaged false positive probabilities of UMR, SSR and $1, 2, 3, 4$-LSR (a) under whitewashing attacks with varying num. of malicious interactions m; and (b) under network dynamics with varying peer lifetime $\frac{1}{\theta}$

For the performance of the LSR system under whitewashing attacks, we only consider the attack strategy that a malicious peer performs a fixed number m of malicious interactions in each session to attack the system for briefness. In Figure 7(a), the time-averaged false positive probabilities for the i-LSR rating systems with $i = 1, 2, 3, 4$, as well as the ones of the UMR and the SSR systems, under varying m from 0 to $1,000$ interactions, are plotted. Again we can see that the i-LSR systems perform between the UMR and the SSR systems, and their performances depend on the system parameter i: the system with a smaller value of i has a better resistance against whitewashing attacks. We also find that the 4-LSR system performs close to the UMR system.

For the influence of network dynamics on the LSR rating system, we obtain the time-averaged false positive probabilities of the i-LSR rating systems, with $i = 1, 2, 3, 4$, and plot them in Figure 7(b). The time-averaged false positive probabilities of the SSR system, as well as the equilibrium false positive probabilities of the UMR system under the same levels of network dynamics are also plotted for comparison. Here we only consider the exponential lifetime model for briefness, and vary the expected peer lifetime from 0 to $1,000$ intervals. It is observed that the i-LSR systems behave differently under network dynamics, with a better failure resilience for the systems with smaller i values. However, unlike the performance under the ideal condition or whitewashing attacks, we find that the UMR system performs better than the 4-LSR system, and it outperforms even the $2, 3$-LSR systems under some dynamics levels. We explain this with the fact that in our UMR rating system, there is a rating recovery mechanism for leaf peers to recover their ratings when their super peers depart.

Our last study for the LSR system focuses on its resistance against false rating attacks. As we have observed in Section 4.4, a strategic false rating attack can influence a rating system in both the confident and the unconfident cases with different attack strategies. For briefness, here we only consider the attack on the confident peers by increasing the relative probability p_l of the colluding malicious peers to get trusted in a peer's credible peer set, and the attack on the unconfident peers by issuing more false ratings f per malicious colluding peer, as these two strategies are proved to be effective in their corresponding cases. The numerical results are plotted in Figure 8. We present the confident case in Figure 8(a) and (b), and plot the unconfident case in Figure 8(c) and (d). We can see from the figures that for both the equilibrium false positive probabilities and the convergence times, the i-LSR systems perform between the UMR and the SSR systems, and a i-LSR system with a larger value of i is more UMR-like with a better attack resistance, while a i-LSR system with a smaller i value behaves more SSR-like, for both the confident and the unconfident cases.

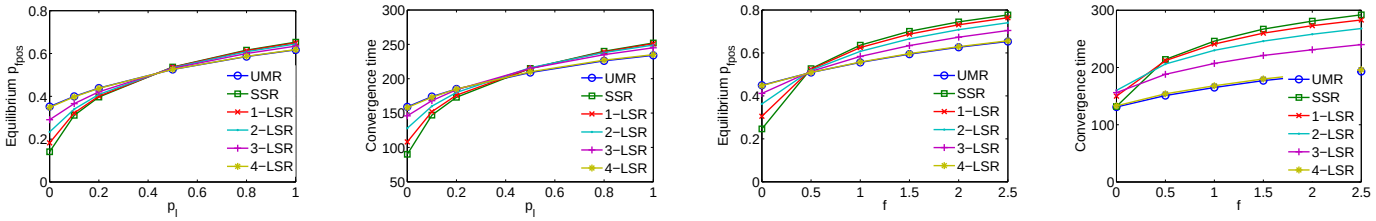
Summarizing all the evaluation results, we conclude that by pre-setting the value of the protocol parameter i properly according to the application context, the LSR rating system could have the desired properties under different adversary attacks and network dynamics. Concretely, a better accuracy, responsiveness, and resistance against whitewashing attacks and network dynamics from the LSR rating system could be expected by setting i small; while a better resistance of the rating system against false rating attacks could be expected by setting i with larger values. Finally, as we have pointed out, the LSR system works inexpensively and scales well with large-sized networks regarding the network overhead.

6. RELATED WORK

The proposal of a distributed rating system on P2P networks was first motivated by eBay's feedback mechanism [1]. In recent years, there are many works on building trust and reputation mechanisms for distributed systems such as P2P networks.

One of the most important components of the quantitative reputation systems is the metrics used for modeling the interaction results. Many systems such as eBay simply use discrete ranks of $[-1, 0, +1]$ for evaluating the interaction results and sum up all the feedbacks as an entity's trust score in reputation. More advanced systems use probabilistic or possibilistic models to express and aggregate the ratings, and to form the belief of trust. For example, the Beta trust model could be found in [13] and [9], the Demspter-Shafer theory based trust model is proposed in [20], and Fuzzy techniques are recently engaged in the system of P2PRep in [21].

For deploying reputation systems on P2P networks, different protocols are proposed for storing and distributing the trust information. As in the design space of P2P lookup systems, P2P reputation systems could be categorized into the unstructured self-managing ones (UMR) and structured supervising



(a) (b) (c) (d)

FIGURE 8. (a) Equilibrium false positive probabilities and (b) convergence times of UMR, SSR, and $1, 2, 3, 4$-LSR systems for confident peers under varying p_l; (c) equilibrium false positive probabilities and (d) convergence times of UMR, SSR, and $1, 2, 3, 4$-LSR systems for unconfident peers under varying f

ones (SSR), and our main focus in this paper is to evaluate and compare the performance of the two designs. For the former category, the P2PRep system [11, 12, 21] designed on Gnutella [2] is a representative example. In P2PRep, similar to Gnutella's file query procedures, a peer asks for ratings on a particular peer by broadcasting a `Poll` message, and peers give their opinions by answering it with a `PollReply` message. Another example for the UMR-like system could be found in [22], in which a two-layered unstructured P2P overlay such as KaZaA [3] is engaged for storing and distributing the trust feedbacks. For the SSR-like system, an early example could be found in [23] which is deployed on a structured P2P overlay called P-Grid [5]. In that system, peers issue evidential feedbacks called complaints and P-Grid is used for distributing and retrieving them by keys. EigenTrust [19] could be viewed as another example of the SSR-like system. In EigenTrust, an inference algorithm similar to Google's PageRank [24] is developed to calculate the global trust score for each peer, and a distributed protocol is designed for mapping a peer to its score manager and calculating the trust scores securely, using DHT techniques such as Chord [4]. The PeerTrust system [10, 25] could also be regarded as a SSR-like system. In PeerTrust, feedbacks on a peer are managed by its supervising peer determined by the P-Grid [5] structured overlay.

For decentralized reputation systems, detecting false feedbacks issued by malicious lying peers is very important. There are many different approaches. In [26], a filtering mechanism is presented to filter out feedbacks with their trust scores falling out of the lower and the higher quantile thresholds. More works focus on building a personalized credibility sub-system for evaluating a peer's trustworthiness in giving feedbacks. In [27], a weighted majority algorithm is used to update a peer's first hand credibility by comparing the interaction result with its recommendation. And [9] proposes a deviation test for the same purpose. The PeerTrust system [10, 25] presents two methods against lying recommenders, namely the TVM and the PSM algorithms. The TVM algorithm simply uses a peer's trustworthiness in providing services as its credibility in giving recommendations, while the PSM algorithm compares the common feedbacks ever issued by the verifying peer and the verified peer, and uses the similarity as the peer's feedback credibility. For solving the sparsity problem caused by the limited number of the credible peers, an inference algorithm is proposed in [28] as a complementary to the PeerTrust system. [29] addresses the issue of separating the service trust and the feedback trust for P2P reputation systems and shows that by decoupling them, the system is more robust against strategic attacks using unfair feedbacks. And in [30], the authors also prove the importance of separately evaluating a peer's credibility in giving opinions in P2P reputation systems.

Besides rating filtering mechanisms, there are other enhancements for P2P reputation systems. An incentive mechanism is proposed in [31] for encouraging rational peers to give honest feedbacks. And [32] points out that by introducing some cost, false feedbacks could be non-profitable for the rational users in an online reputation system. In [33], a proactive reputation system is proposed for speeding up the procedure of accumulating feedbacks by proactively probing the target peer. [8] discusses possible methods for preventing Sybil attack and in [34], a social network based random walk algorithm is presented for limiting the number of Sybil peers a loyal peer may encounter in a voting-like procedure (e.g. reputation).

However, despite the abundance of solutions in this area, systematic performance evaluations, especially theoretical ones are relatively rare. [35] gives a simulation-based investigation on the performance of an UMR-like reputation system, and shows that the system could be more effective with limited sharing of the ratings even under 40% of attackers in the network. By contrast, we develop our studies based on an analytical modeling framework which covers the UMR-like and the SSR-like rating systems, therefore providing a broad and profound insight on the performance of P2P rating systems.

7. CONCLUSION

In this paper, we have compared the accuracy, responsiveness, and resistance under network dynamics as well as various adversary attacks for the two representative P2P rating systems, namely the unstructured self-managing rating (UMR) system and the structured supervising rating (SSR) system. A stochastic analytical model is proposed for this purpose. We present theoretical studies for both systems' performances under different network environments with varying intensities of peer failures and adversary attacks, and find that both designs have merits as well as flaws. A configurable loosely supervising rating (LSR) system has been proposed in this paper, and we show via numerical analysis that with a proper pre-setting of a protocol parameter, the LSR system could tradeoff the merits of both the UMR and the SSR systems, and yield a desired overall performance under a specific application context.

REFERENCES

[1] Resnick, P., Zeckhauser, R., Swanson, J., and Lockwood, K. (2006) The value of reputation on ebay: a controlled experiment. *Experimental Economics*, **9**, 79–101.

[2] Gnutella. http://www.gnutella.com.

[3] Kazaa. http://www.kazaa.com.

[4] Stoica, I., Morris, R., Liben-Nowell, D., Karger, D. R., Kaashoek, M. F., Dabek, F., and Balakrishnan, H. (2001) Chord: A scalable peer-to-peer lookup service for internet applications. *Proceedings of SIGCOMM 01*, San Diego, CA, 27-31 August, pp. 149–160. ACM Press, New York, NY.

[5] Aberer, K., Cudre-Mauroux, P., Datta, A., Despotovic, Z., Hauswirth, M., Punceva, M., and Schmidt, R. (2003) P-grid: A self-organizing structured p2p system. *ACM SIGMOD Record*, **32**, 29–33.

[6] Papaioannou, T. G. and Stamoulis, G. D. (2004) Effective use of reputation in peer-to-peer environments. *Proceedings of CCGrid 04*, Chicago, IL, 19-22 April, pp. 259–268. IEEE Computer Society, Washington, DC.

[7] Dumitriu, D., Knightly, E., Kuzmanovic, A., Stoica, I., and Zwaenepoel, W. (2005) Denial-of-service resilience in peer-to-peer file sharing systems. *Proceedings of SIGMETRICS 05*, Banff, Canada, 6-10 June, pp. 38–49. ACM Press, New York, NY.

[8] Douceur, J. R. (2002) The sybil attack. *LNCS 2429: Proceeding of IPTPS 02*, Cambridge, MA, 7-9 March, pp. 251–260. Springer-Verlag, Berlin.

[9] Buchegger, S. and Le Boudec, J. (2004) A robust reputation system for p2p and mobile ad-hoc networks. *Proceedings of P2PEcon 04*, Cambridge, MA, 4-5 June. Harvard University, Cambridge, MA.

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

[11] Damiani, E., di Vimercati, D., Paraboschi, S., Samarati, P., and Violante, F. (2002) A reputation-based approach for choosing reliable resources in peer-to-peer networks. *Proceedings of CCS 02*, Washington, DC, 18-22 November, pp. 207–216. ACM Press, New York, NY.

[12] Damiani, E., di Vimercati, D., Paraboschi, S., and Samarati, P. (2003) Managing and sharing servents' reputations in p2p systems. *IEEE Trans. on Data and Knowledge Engineering*, **15**, 840–854.

[13] Jøsang, S., Hird, S., and Faccer, E. (2003) Simulating the effect of reputation systems on e-markets. *LNCS 2692: Proceedings of iTrust 03*, Crete, Greece, 28-30 May, pp. 179–194. Springer-Verlag, Berlin.

[14] LCA-REPORT-2005-013 (2004) *Understand the simulation of mobility models with Palm calculus.* EPFL. Lausanne, Switzerland.

[15] Trivedi, K. S. (2002) *Probability and Statistics with Reliability, Queueing and Computer Science Applications, 2nd Edition.* John Wiley & Sons, New York, NY.

[16] Leonard, D., Rai, V., and Loguinov, D. (2005) On lifetime-based node failure and stochastic resilience of decentralized peer-to-peer networks. *Proceedings of SIGMETRICS 05*, Banff, Canada, 6-10 June, pp. 26–37. ACM Press, New York, NY.

[17] Saroiu, S., Gummadi, P. K., and Gribble, S. D. (2002) A measurement study of peer-to-peer file sharing systems. *Proceedings of MMCN 02*, San Jose, CA, 18-25 January, pp. 156–170. SPIE Press, Bellingham, WA.

[18] Stutzbach, D. and Rejaie, R. (2005) Characterizing the two-tier gnutella topology. *Proceedings of SIGMETRICS 05*, Banff, Canada, 6-10 June, pp. 402–403. ACM Press, New York, NY.

[19] Kamvar, S. D., Schlosser, M. T., and Garcia-Molina, H. (2003) The eigentrust algorithm for reputation management in p2p networks. *Proceedings of WWW 03*, Budapest, Hungary, 20-24 May, pp. 310–317. ACM Press, New York, NY.

[20] Yu, B. and Singh, M. P. (2002) An evidential model of distributed reputation management. *Proceedings of AAMAS 02*, Bologna, Italy, 15-10 June, pp. 294–301. ACM Press, New York, NY.

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

[22] Mekouar, L., Iraqi, Y., and Boutaba, R. (2005) Peer-to-peer's most wanted: malicious peers. *Computer Networks*, **50**, 545–562.

[23] Aberer, K. and Despotovic, Z. (2001) Managing trust in a peer-to-peer information system. *Proceedings of CIKM 01*, Atlanta, GA, 5-10 October, pp. 310–317. ACM Press, New York, NY.

[24] SIDL-WP-1999-0120 (1999) *The pagerank citation ranking: bringing order to the web.* Stanford University. Stanford, CA.

[25] Srivatsa, M., Xiong, L., and Liu, L. (2005) Trustguard: countering vulnerabilities in reputation management for decentralized overlay networks. *Proceedings of WWW 05*, Chiba, Japan, 10-14 May, pp. 422–431. ACM Press, New York, NY.

[26] Whitby, A., Jøsang, A., and Indulska, J. (2005) Filtering out unfair ratings in bayesian reputation

systems. *The Icfaian Journal of Management Research*, **4**, 48–64.

[27] Yu, B. and Singh, M. P. (2003) Detecting deception in reputation management. *Proceedings of AAMAS 03*, Melbourne, Australia, 14-18 July, pp. 73–80. ACM Press, New York, NY.

[28] TR-2005-017-A (2005) *Countering sparsity and vulnerabilities in reputation systems.* Emory University. Atlanta, GA.

[29] Swamynathan, G., Zhao, B. Y., and Almeroth, K. (2005) Decoupling service and feedback trust in a peer-to-peer reputation system. *LNCS 3759: Proceedings of AEPP 05*, Nanjing, China, 2-4 November, pp. 82–90. Springer-Verlag, Berlin.

[30] Selcuk, A. A., Uzun, E., and Pariente, M. R. (2004) A reputation-based trust management system for p2p networks. *Proceedings of CCGrid 04*, Chicago, IL, 19-22 April, pp. 251–258. IEEE Computer Society, Washington, DC.

[31] Papaioannou, T. G. and Stamoulis, G. D. (2005) An incentives' mechanism promoting truthful feedback in peer-to-peer systems. *Proceedings of CCGrid 05*, Cardiff, UK, 9-12 May, pp. 275–283. IEEE Computer Society, Washington, DC.

[32] Bhattacharjee, R. and Goel, A. (2005) Avoiding ballot stuffing in ebay-like reputation systems. *Proceeding of P2PEcon 05*, Cardiff, UK, 22-22 August, pp. 133–137. ACM Press, New York, NY.

[33] Swamynathan, G., Zhao, B. Y., and Almeroth, K. (2006) Exploring the feasibility of proactive reputations. *Proceedings of IPTPS 06*, Santa Barbara, CA, 26-28 February. University of California - Santa Barbara, Santa Barbara, CA.

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

[35] Marti, S. and Garcia-Molina, H. (2004) Limited reputation sharing in p2p systems. *Proceedings of EC 04*, New York, NY, 17-20 May, pp. 91–101. ACM Press, New York, NY.