From Hazardous Behaviours to a Risk Metric for Reputation Systems in Peer to Peer Networks

Erika Rosas, Xavier Bonnaire

From Hazardous Behaviours to a Risk Metric for Reputation Systems in Peer to Peer Networks

Erika Rosas[1] and Xavier Bonnaire[2]

[1] Laboratoire d'Informatique de Paris 6
Université Pierre et Marie Curie - CNRS UMR 7606
Paris - France
[2] Universidad Técnica Federico Santa María
Departamento de Informática
Valparaíso - CHILE

Abstract. Peer to Peer systems have shown to be very powerful to build very large scale distributed information systems. They are self organized, and provide very high availability of the data. However, the management of malicious peers is a very open problem for the Peer to Peer research community, and building trust is a very difficult task.
In this context, Reputation Systems have shown to be a very good solution to build trust in Peer to Peer systems. Nevertheless, using only the reputation value of a peer to decide to make a transaction is not sufficient to guarantee that it will succeed, and the use of the credibility of recommendation emitters does not always significantly mitigate the computed reputation.
We show in this paper the importance of the notion of risk associated to the reputation value, and why a better decision can be taken using both, the reputation and a risk value, for a given peer. We present some metrics based on the list of recommendations for a peer that allow to detect some suspicious behaviours that can alert the application of the presence of a malicious peer. The proposed metric is flexible such that an application can adapt the metric to its needs, given more or less weight to some specific types of behaviours.
We present some simulations to show the influence of malicious behaviours of a peer over its reputation value with the evaluation of the associated risk, and how our metric can detect this kind of behaviours. We conclude about the need to use a risk factor associated to the reputation value, and present some future works about the risk metrics.

1 Introduction

Building trust in Peer to Peer networks is a very difficult task, mainly because of the number of peers, the high dynamism of the network, and the presence of malicious peers. These characteristics make using a certification authority based on a set of servers not a very well suited answer to this problem, as it requires a central administration, which it is not a scalable solution. Other traditional authentication techniques cannot be used because of the ability of a peer to change its identity, and the need of anonymity of the peers [1].

In this context, reputation systems have shown to be a very good solution to build trust in Peer to Peer systems [12], [6], [14], [9], [10]. The key idea of a reputation system is to provide a reputation value for each peer, which can be seen as the probability for the peer to be trusted. To compute the reputation value the system defines a *metric* based on a set of recommendations emitted by other peers after completing a transaction. When a transaction succeeds, a good recommendation must be emitted, and a bad one otherwise. An application can then decide whether or not to do a transaction with a peer according to its reputation value.

Usually, the metric of reputation systems also considers the credibility of the peer which emits the recommendation, as a function of its reputation value [13] [9] [6] or as the similarity of its past evaluations [12] [2]. Nevertheless, the reputation value is not sufficient, and malicious peers can take advantage of a good reputation value to deceive other peers.

As the reputation value is based on the behaviour of the peers, it cannot reflect some of the strategies used by the peers to fool the reputation system. This is why the notion of risk has been introduced as a complement to the reputation value. The risk value is used to try to detect suspicious behaviours of the peers that have a good reputation and seem to be trusted.

To our knowledge, the notion of risk as presented in this paper has never been proposed before. Only the work in the Pet [8] reputation system introduces the notion of risk in their trust model. In Pet, this is a value derived from direct interactions with other peers. This is a very different approach since it only take into account a short-term behaviour [8], and it is focused to detect sudden changes of behaviour of the peers that the reputation value cannot detect. A drawback of this work is that in peer to peer networks, with millions of peers it is not very probable that a peer had already a previous direct interaction with other specific one.

A few proposals have attempted to address the issue of malicious attacks to the reputation system. Overall, reputation system are focused on mitigating malicious recommendations, which are detected with the use of a credibility value. Xiong and Liu in [12] consider the problem of free riders adding to the reputation metric a community context factor, which can be a function of the feedback provided by the peer to the reputation system. This is a way to encourage the participation of peers.

TrustGuard [11] is a framework that is focused, as our work, on understanding the vulnerabilities of the reputation systems and on how to minimize the effects of malicious peers. The difference is that TrustGuard changes the reputation metric to achieve this. We believe that the reputation metric gives valuable information itself and can be quite flexible for an applications, but we also believe that an application needs additional information to know if the reputation value of a peer can be trusted itself. TrustGuard [11] detects three vulnerabilities, malicious peers that adapt its behaviour to maximize its malicious goals, rumors and false recommendations. We did not consider in our work the last two problems because they can be mitigated directly in the reputation metric. A solution based on a proof of transaction (evidence) has been proposed in [11]. We will see later on in this paper that our approach of the risk uses an analysis of the behaviour of the peer, based on the list of recommendations that the reputation system already has to calculate the reputation value.

The RQC reputation system [5] proposes a quality function to evaluate the trustworthiness of the reputation value. Similarly to some metric in our work, they consider the number of recommendations and the variance of the data to compute the quality of the reputation value. RQC searches the consistence in the reputation value more that to detect suspicious attack of malicious peers that take advantage of their reputation value to attack the system.

In this paper, we propose a risk metric capable to detect several well-known malicious behaviours of peers, such that the Oscillating Personality, the Random Behaviour, and the Repeated One Shot Attack.

The rest of the paper is organized as follows. In Sect. 2, we briefly present a general model of a reputation system where a risk metric can be applied. Section 3 details a set of risk metrics to detect several well-known malicious behaviours of peers. Then, experiments and results are shown in Sect. 5. Finally, conclusion and future work are presented in Sect. 6.

2 Reputation System Model

The risk metrics presented in the next section are based in the idea that to compute the reputation value of a peer X ($Re(X)$) the reputation system collects a number of recommendation emitted by some peers which already had transactions with X in the past.

We note $F_i(X)$ the recommendation emitted about a peer X of index i from a total of m recommendations. The value m in some systems can be considered like a sufficient number of recommendations or in others as the maximal number of recommendations to compute a reputation.

We suppose in the following that the reputation value is the probability for a peer to be trusted, and that the reputation system uses recommendations in the range [0..1], with at least three discret values.

There are several reputation systems that follow this model [9], [12], [2], [13]. All of them could include a risk metric as a complement to the reputation value in order to help an application to decide whether or not to make a transaction.

3 Malicious Behaviours and Associated Risk Metrics

There are several strategies that a malicious peer can use to fool the reputation system. None of them can be detected using only the reputation value of the peer. An application can then ignore a wrong behaviour of this peer. In this section, we present a set of well-known malicious behaviours for a peer, and we propose an associated risk metric capable of detecting this malicious behaviour.

3.1 White Washers

A peer is called a White Washer when it intentionally leaves the network and enter again with a new identity, in order to clear its history of recommendations. This allows the peer to fool an application, appearing with a fresh good reputation. This is mainly due to the assignment of a good reputation to new peers entering the network (positive discrimination) to give them a chance to make a transaction. Therefore, it becomes difficult to discriminate new peers from malicious ones for the reputation system. The worst case appears in the Sybil Attack [3] where a peer can have multiple identities.

In decentralized reputation systems there are no solutions to identify these peers, but there are some ways to mitigate their impact. The use of expensive identifiers can help to prevent a peer from trying to get several different identifiers, due to the computational or financial cost to obtain a new identifier.

Giving a reputation to the resources used in the network (i.e. files, etc...) like in [2] [7], or giving a low reputation value to new peers can help to mitigate the effects of White Washers. However, this does not encourage new honest peers to participate to the system. The work of Friedman in [4] has shown that the distrust in new peers is a social cost inherent to the easy change of identity.

The problem with the reputation value is that a peer X with a number of good recommendations $r \ll m$, will have a similar reputation value that a peer with m good recommendations. For example, a new peer with only one good recommendation will have nearly the same reputation value of a peer with m good recommendations.

To mitigate the effect of White Washers, we propose the risk metric given by (1), where r is the number of recommendations that have been emitted about peer X.

$$Ri_A(X) = \left(1 - \frac{r}{m}\right) \tag{1}$$

The result is a number in the range $[0, 1]$, 0 means no risk, the peer has a sufficient history of recommendations and the reputation value can be taken into account without risk. On the other hand, a risk of 1 means that the reputation value is very risky because there is not enough information about X, and the computed value is the default for new peers.

3.2 Oscillating Personality

The problem of oscillating personality appears because the reputation value is generally an average or a weighted average of the recommendations that have been emitted about a peer. The result gives a global idea of the past behaviour of the peer.

A peer which makes a good transaction and a bad one in turn will have a reputation value in the middle range, and can be seen like a peer that has an average behaviour. However, this peer is a malicious peer that makes good recommendations to balance its bad behaviour and to continue appearing like an average peer, instead of a malicious one. It can be more interesting for an application to choose a peer with a more regular behaviour than a very irregular one.

We use the standard deviation of the emitted recommendations to detect this kind of behaviour. The bigger is the standard deviation, the farther are the recommendation from the average. A value of 1 means that there is a risk of 100%, and 0 means no risk (i.e. all the recommendations are near to the average value).

The metric in (2) allows to detect an oscillating personality. The role of factor 4 is to normalize the equation to obtain a value in $[0, 1]$, r is the number of recommendations used to compute the risk, and $F_i(X)$ is the recommendation of index i about peer X.

$$Ri_B(X) = 4 \times \frac{\sum_{i=0}^{r}(F_i(X) - \overline{F(X)})^2}{k} \tag{2}$$

3.3 Random Behaviour

A peer has a random behaviour when the recommendations emitted for this peer are fully distributed in the range of possible recommendations (in our case in the range [0..1]). A Byzantine peer can have this kind of behaviour. From the reputation system point of view, this type of peers will have the same reputation value than ones with a permanent regular behaviour.

This is significantly different from the previous case because for a random behaviour, the standard deviation of the emitted recommendations for this peer will not result in a high value.

Thus, we use the entropy of the recommendations values to detect this type of behaviour. The entropy is an indicator of the level of disorder in the data. A peer with low entropy is a peer with no disorder in the recommendations, which means that its behaviour has always been the same. A peer with a high entropy, is a peer with recommendations values fully dispersed in the range of recommendation.

$$Ri_C(X) = \frac{\sum_{j=1}^{l} p_X(x_j)\log_2(p_X(x_j))}{\log_2(l)} \tag{3}$$

Equation 3 shows the risk metric to detect this kind of behaviour, where l is the number of possible values for a recommendation (cardinality of the set of discrete recommendation values), and $p_X(x_1)$ is the number of recommendation with the value x_1 for X divided by the total number of recommendations.

For a reputation system with a continuous range of recommendation values, for example [12] in the range $[0, 1]$, applying this metric requires to make the range discrete. An example of discretization can be can be that the range $[0, 0.2]$

is assigned to $p_X(x_1)$, that is, all the values in that range counts to compute the probability $p_X(x_1)$.

The denominator of (3) is a normalization factor. The result is in the range $[0, 1]$. The numerator represents the maximal possible entropy with all the values equally dispersed in the l possible categories of the recommendation values.

3.4 Repeated One Shot Attack

A One Shot Attack occurs when a peer, which is apparently a good one, makes sparse bad transactions. As most of the transactions of the peer are good ones, the bad transactions do not make significant changes to the overall reputation of the peer that will be a good reputation. This is absolutely impossible to detect for an application, using only the reputation value.

In the reputation system proposed in [9], a behaviour like the one illustrated in Fig. 1 gives a reputation value of 0.8 (considering equal credibility values for all the evaluators). This value does not show that this peer is a malicious peer which has a malicious behaviour every 3 transactions.

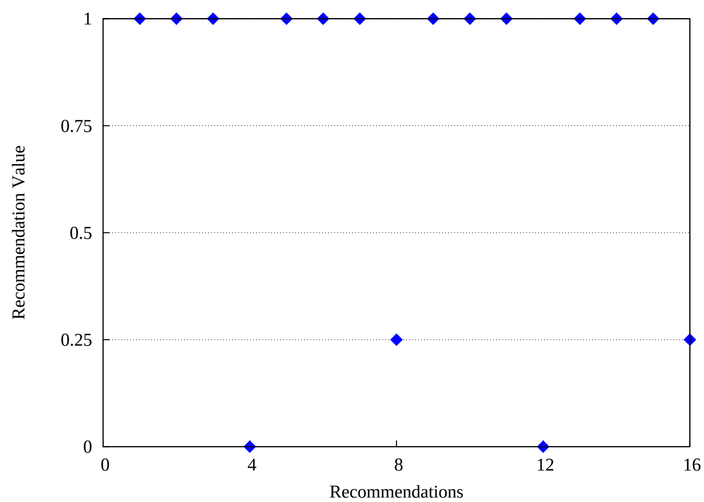

Fig. 1. Example repeated one time attack

The risk metric we propose to detect the Repeated One Shot Attack is based on the analysis of the difference among consecutive recommendation values for a peer. The attack is only possible if the recommendations are clearly partitioned into two groups, with good and bad recommendations (there is no average recommendation), and if there are only sparse bad ones. In this case, the risk metric propose in (4) gives an evaluation of the risk, and 0 otherwise.

Only when there are more stable and good recommendations than bad ones there is a possibility of this attack, for this reason (5) gives 0 risk otherwise.

A recommendation value will be considered suspicious if the difference between itself and the previous transaction is bigger than a value D, that depends on the range of the recommendation values. A value of D equal or bigger to $0,5$ would be a adapted difference in a recommendation value range of $[0,1]$. In (5) r is the number of recommendation the system has about X.

$$J(X,i) = \begin{cases} 1 \quad if \quad |F_i(X) - F_{i-1}(X)| < D \\ 0 \; oherwise \end{cases} \tag{4}$$

$$Ri_D = \begin{cases} \dfrac{\sum\limits_{i=1}^{r} J(i,X)}{r - \sum\limits_{i=1}^{r} J(i,X)} & if \quad \sum_{i=1}^{r} J(i,X) < \frac{r}{2} \\ 0 & otherwise \end{cases} \tag{5}$$

4 Global Metric

We have presented a set of risk metrics to help an application in the decision process to make a transaction with a given peer. A global risk can be computed according to the applications needs. The factors alpha, beta, gamma and delta allow the application to give more weight to each term according to its requirement. Equation 6 gives the global risk computation.

$$Ri_{Global}(X) = \frac{\alpha Ri_A(X) + \beta Ri_B(X) + \gamma Ri_C(X) + \delta Ri_D(X)}{\alpha + \beta + \gamma + \delta} \tag{6}$$

The sum of all factors is used to maintain the result within the range [0..1].

To decide whether or not making a transaction with a given peer X, an application has two indicators, the reputation $Re(X)$ of peer X, and the global risk value $Ri_{Global}(X)$ associated to X. The use of the reputation value and the risk value completely depends on the application needs.

The reputation value of a peer with a low risk means that the reputation value effectively reflects the past behaviour of the peer. A high risk means that the reputation value does not necessarily reflects the past behaviour of the peer, and making a transaction with this peer may be hazardous. Nevertheless, a high risk does not means that the peer is a malicious one, it is only a high probability, and the transaction may succeed.

For completeness, it is worth to mention that there are two other types of malicious behaviours that were not considered in this work: milking personality and false recommendations. The reason is because they can be easily detected during the reputation value calculation.

Milking personality is the strategy of a peer that builds a good reputation value and after some time starts having a bad behaviour. As its reputation value is high, the peer can deceive other peers until its reputation value will fall. To detect this behaviour the metric for the reputation value can add a fading factor, which gives more weight to the latest recommendations. False recommendations are the recommendations emitted by malicious peers about other peers, but they do not reflect the peer's behaviour during the transaction. The system can use a credibility value to detect this behaviour.

In the next section, we present some simulation results to show the efficiency of our metric.

5 Results and Analysis

The experiments have been done in order to quantify the efficiency of the risk metrics front of the correspondent attack. All of them have been done using the reputation system proposed in [9]. This reputation system uses a list of the last m recommendations emitted about a peer to compute its reputation value. In the experiments the size of the recommendation list has been set to $m = 16$, because this value has shown to be the best choice for this reputation system (See [9]).

In all the experiments, the total number of peers is $100,000$, which make approximately 100 transactions each. The results are averaged every $200,000$ transactions. For each transaction, a peer A randomly chooses a peer B in the network to make the transaction. To decide whether or not to make the transaction the risk and reputation value are aggregated using (7). This value is used as a threshold to probabilistically decide to accept or deny the transaction. The key idea in (7) is to increase or decrease the threshold according to the reputation and risk values.

$$Th_t(B) = \begin{cases} \text{If} & 0.75 < Re_t(B) \leq 1 & Re_t(B) \times \left(1 - \frac{Ri_t(B)}{2}\right) \\ \text{If} & 0.25 \leq Re_t(B) \leq 0.75 & Re_t(B) \times (1 - Ri_t(B)) \\ \text{If} & Re_t(B) \leq 0.25 & Re_t(B) \times (1 + 2 \times Ri_t(B)) \end{cases} \quad (7)$$

The first experiment is about White Washers. 20% of peers in the system are White Washers. They make malicious transactions and when their reputation value drops down to 0.05 they leave the system and join again with a clean new identity.

Figure 2 shows the accepted transactions to white washer in the reputation system with the risk metric and without it. Malicious transactions decrease in more than a 40%.

In this case, the risk metric affects the new honest peers in the system, but as they continue to do honest transactions to obtain good recommendations, the risk value rapidly falls to 0 and stops affecting the transactions between these peers. Figure 3 shows the evolution on the risk value for honest peers and for the malicious ones. The results represents the average of the risk of the set of peers. We see in this graphic that the risk for the honest peers goes down as they make more transactions in the system.

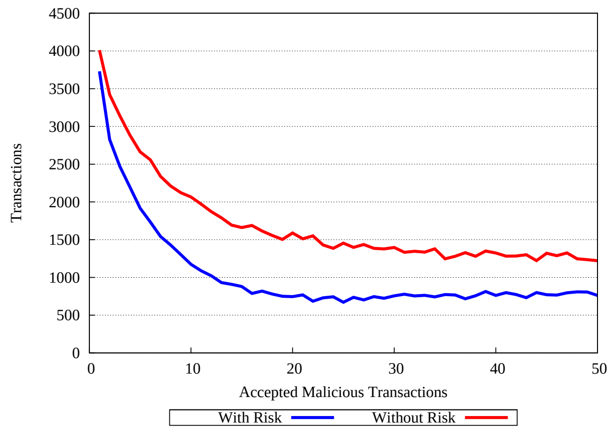

Fig. 2. Accepted Transactions to White Washers

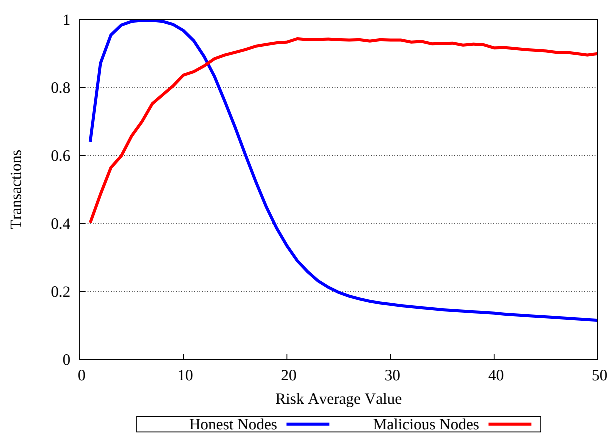

Fig. 3. Risk Value evolution with White Washers

The second type of behaviour to analyze is the oscillating personality. In this experiment we have considered malicious peers that make a good and a bad transaction in turn to continue with a regular reputation value. The results are showed in Fig. 4. The accepted malicious transaction drop in more than 80% which shows that our metric is very efficient to detect this type of behaviour. In this case, honest peers are minimally affected by the risk metric since they usually make good recommendations. Moreover, the number of false recommendations is not sufficient to get a high risk.

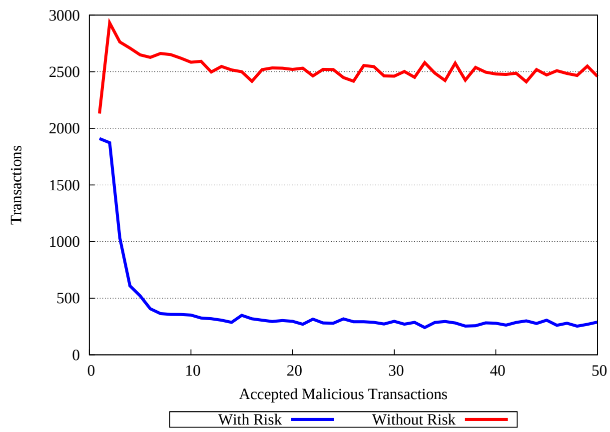

Fig. 4. Accepted Transactions to Oscillating Personality

The results for the analysis of the metric presented for the random behaviour are presented in Fig. 5. This figure shows that without the risk metric, 20% of the malicious peers make 1500 bad transactions. Using the risk metric based on the entropy, the number of malicious transactions falls under 250, which represents an improvement of more than 80%.

The last experiments analyze the behaviour of the metric in front of the Repeated One Shot Attack. The parameter D have been set to 0.5 which is half of the total range. In this case, we have considered malicious peers that repeatedly make 3 good transactions and then a bad one. The results are shown in Fig. 6.

This metric avoids making around 40% of malicious transactions. Honest peers are only affected by this metric if there are false recommendations in the system. If there is a high percentage of lying peers, the metric could think this is a Repeated One Shot Attack. This really depends on how long is the list of recommendations considered in the risk and reputation computation

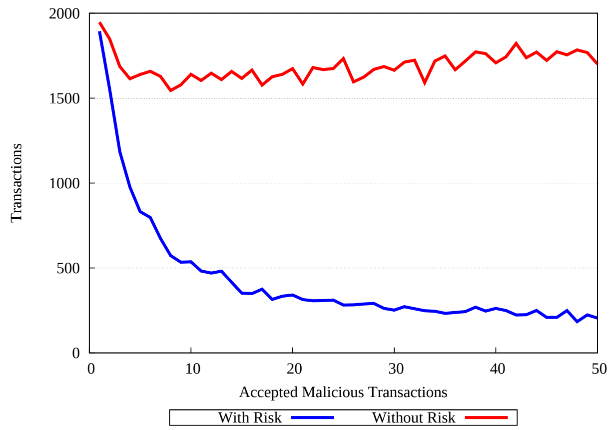

Fig. 5. Accepted Transactions to Random Behaviour

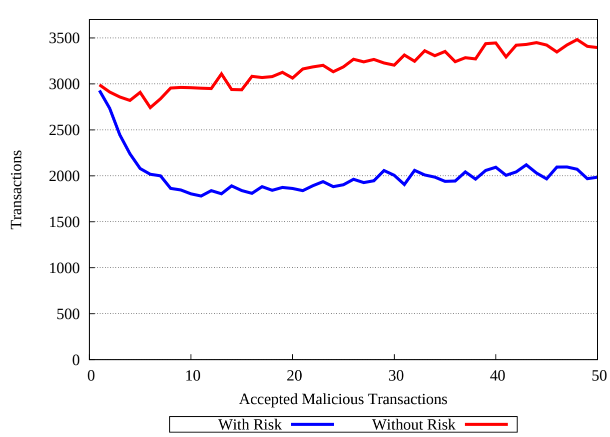

Fig. 6. Accepted Transactions to Repeated One Time Attack

6 Conclusion

This works introduces the concept of risk metric in reputation systems to complement the reputation value and to detect some suspicious behaviour ignored by the reputation value. We have presented four risk metrics based on the analysis of the list of recommendation the reputation system has about a given peer.

The experiments have shown very good results in the detection of the attacks and a clear fall in the number of malicious transaction made by peers with wrong behaviour (up to an 80%). The risk that has been proposed helps to trust the reputation value itself, preventing an application from making very hazardous transactions.

Further efforts have to be made to detect other kinds of attacks to reputations systems. We especially think about the detection of White Washers which is a difficult task for reputation systems.

Further work also consists in creating risk metrics for other types of reputation system, like the ones based on transitive reputation. Another pending issue is to test different aggregation schemes for the risk and the reputation value, depending on the requirements of the application.

References

1. Xavier Bonnaire and Erika Rosas. A critical analysis of latest advances in building trusted p2p networks using reputation systems. In *WISE2007 Workshops: The 8th International Conference on Web Information Systems Engineering*, Lecture Notes in Computer Science 4832, pages 130–141. Springer-Verlag, December 2007. (To appear).
2. Ernesto Damiani, Sabrina De Capitani di Vimercati, Stefano Paraboschi, and Pierangela Samarati. Managing and sharing servents' reputations in p2p systems. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):840–854, July/August 2003.
3. John Douceur. The sybil attack. In *IPTPS '02: Proceedings of the First International Workshop on Peer-to-Peer Systems*, volume 2429 of *Lecture Notes in Computer Science*, pages 251–260, Cambridge, MA, USA, March 2002. Springer.
4. Eric Friedman and Paul Resnick. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, 10(2):173–199, June 2001.
5. Anurag Garg, Anurag Garg, Roberto Battiti, Roberto Battiti, Gianni Costanzi, and Gianni Costanzi. Dynamic self-management of autonomic systems: The reputation, quality and credibility (rqc) scheme. In *In The 1st IFIP TC6 WG6.6 International Workshop on Autonomic Communication (WAC*. Springer-Verlag, 2004.
6. Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web*, pages 640–651, New York, NY, USA, 2003. ACM Press.
7. So Y. Lee, O-Hoon Kwon, Jong Kim, and Sung J. Hong. A reputation management system in structured peer-to-peer networks. In *WETICE '05: Proceedings of the 14th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise*, pages 362–367, Washington, DC, USA, June 2005. IEEE Computer Society.

8. Zhengqiang Liang and Weisong Shi. Pet: A personalized trust model with reputation and risk evaluation for p2p resource sharing. In *HICSS '05: Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, page 201.2, Washington, DC, USA, January 2005. IEEE Computer Society.
9. Erika Rosas. Diseño e implementación de un sistema de reputation para redes p2p mixtas. Master's thesis, Universidad Técnica Federico Santa María, November 2007.
10. Aameek Singh and Ling Liu. Trustme: Anonymous management of trust relationships in decentralized p2p. In *P2P '03: Proceedings of the IEEE International Conference on Peer-to-Peer Computing*, page 142, Washington, DC, USA, September 2003. IEEE Computer Society.
11. Mudhakar Srivatsa, Li Xiong, and Ling Liu. Trustguard: countering vulnerabilities in reputation management for decentralized overlay networks. In *WWW '05: Proceedings of the 14th international conference on World Wide Web*, pages 422–431, New York, NY, USA, 2005. ACM Press.
12. Li Xiong and Ling Liu. Peertrust: supporting reputation-based trust for peer-to-peer electronic communities. *IEEE Transactions on Knowledge and Data Engineering*, 16(7):843–857, 2004.
13. Bin Yu, Munindar Singh, and Katia Sycara. Developing trust in large-scale peer-to-peer systems. In *MAS&S2004: Proceedings og the IEEE First Symposium on Multi-Agent Security and Survivability*, pages 1–10, Philadelphia, Pennsylvania, USA, August 2004. IEEE.
14. Runfang Zhou and Fellow-Kai Hwang. Powertrust: A robust and scalable reputation system for trusted peer-to-peer computing. *IEEE Transactions on Parallel and Distributed Systems*, 18(4):460–473, April 2007.