Send Orders for Reprints to reprints@benthamscience.ae

The Open Cybernetics & Systemics Journal, 2015, 9, 1393-1398

Open Access

A New Reputation Model for P2P Network Based on Set Pair Analysis

He Chaokai[1,*] and Wu Meng[2]

[1]School of Computer science, Nanjing University of Posts & Telecommunications, Nanjing, Jiangsu, 210003, P.R. China; [2]College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, 210003, P.R. China

Abstract: P2P reputation systems are useful to evaluate the trustworthiness of peers and to combat malicious and selfish peer behaviors. The reputation system assigns each peer a global reputation score through collecting locally generated peer feedbacks and aggregation. In traditional reputation system, a peer rates the other by a score, and after a transaction is finished, no more detail of transaction is considered in the process. In order to solve this problem, a novel reputation aggregation scheme is proposed due to which detail of transaction could be considered, based on set pair analysis (SPA). The empirical evaluation results reveal that the proposed approach is scalable, accurate, robust, and fault-tolerant.

Keywords: P2P network, Reputations, Set pair analysis, Trust.

1. INTRODUCTION

P2P networks are networks in which peers can freely join in and leave the systems. Each peer is both the provider of resources and consumer and can access each other directly. P2P networks have many benefits over client-server approaches to file distribution, including robustness, scalability, and diversity of available data. Due to openness, anonymity, and dynamic nature of peer activities, P2P networks are very vulnerable and are easily abused by selfish and malicious peers [1]. Some peers perform malicious behaviors which include providing false services, spreading malicious viruses, and so forth.

In order to encourage peers to participate in the network and combat malicious peers, reputation system takes an important role by distinguishing different peers according to their historical behavior [2]. It is obvious that peers which provide reliable services act or declare to have higher reputation value. With an efficient reputation system, peers do not hesitate to interact with unknown peers. Furthermore, in commercial P2P applications, such as P2P auctions, pay-per transaction, trusted content delivery, and P2P service discovery, there is a great demand to identify trustworthy peers. With the evolution and acceptance of these P2P services, p2p network has started putting more demand on the efficiency and accuracy of the online reputations system.

In general, after a transaction is completed, the participating peer rates the provider based on its experience in the process. The reputation system computes each peer's global reputation score by aggregating the local rates from those peers which interacted. Before a transaction, a peer launches a request by accessing the global reputation scores of the respondents and peers are able to choose the respondent with high reputation value to finish the transaction.

*Address correspondence to this author at the school of computer Science Nanjing University of Posts & Telecommunications, Nanjing, Jiangsu, 210003, P.R. China; Tel: +86 25 858665258; Fax: +86 25 85866258; E-mail: hck@njupt.edu.cn

The field of P2P reputation systems has also commanded increasing attention. Reputation aggregation is the most important issue involved in this process, the function of which is to yield global reputation scores from locally generated feedbacks. PowerTrust [3], EigenTrust [4], PeerTrust [5], P-Grid [6], TrustMe [7], GoosipTrust [8] *etc*. are the most popular models.

Taking advantage of the power law distribution of peer feedbacks, the PowerTrust can aggregate global reputations fast. Self-policing, anonymity, no profit to new comers, minimal overhead, robust to malicious collectives *etc*. are five issues that are important to address in the P2P reputation system proposed by The EigenTrust algorithm, which presents a distributed and secure method to compute peers' global reputation values, based on Power iteration. These five factors include feedback in terms of amount of satisfaction, numbers of transaction, credibility of feedback, transaction context factor, and community context factor. PeerTrust computes the trustworthiness of a given peer. PGrid is the first trust management study for unstructured P2P networks, which is based on a decentralized storage method. By using a random assignment of reputation-holding peers and employing smart Public Key mechanisms to keep the anonymity, TrustMe performances are based on trusting the management. By resorting to gossip-based protocol and identity-based cryptography, GoosipTrust is adapted to peer dynamics and robust to disturbance.

The remainder of this paper is organized as follows. Section 2 first reviews the architecture and workflow of reputation system. The theory of set pair analysis is presented in section 3. The detail of the new algorithm is discussed in section 4. Empirical results and analysis thereof are given in section 5, while section 6 concludes the paper.

2. ARCHITECTURE AND WORKFLOW OF REPUTATION SYSTEM

In general, a reputation system mainly consists of three parts; 1) Information gathering: it refers to collecting the

1874-110X/15

information on the past transactional behavior of each peer which is also the basis of the reputation system. 2) Aggregating: it is carried out according to the information collected and reputation system scores and the peers are ranked on its basis. 3) Punishment and reward: reputation system takes action against malicious peers and rewarding contributors. Each component requires separate system mechanisms as shown in Fig. (**1**) [9].

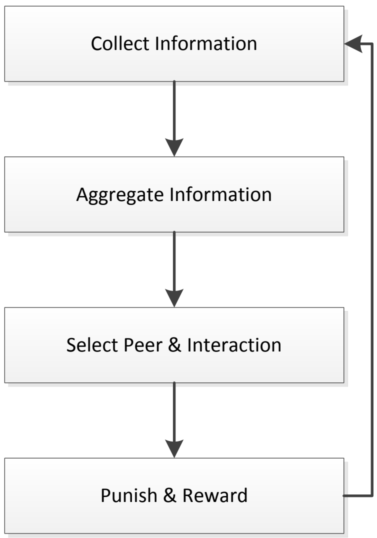

Fig. (1). The work flow of reputation systems.

In P2P network, the completion of a transaction mainly needs to go through three steps: send a request, aggregate received message, make a decision. Each peer's own reputation value is stored by a set of triples Ψ={ID, num.S, num.F}, which can be communicated with others, as established by the P2P communication protocol, also a self-appointed ID. Num.S refers to the number of successful transactions and num.F refers to the number of failed transactions.

According to the above information, peer needs a suitable algorithm to complete the transformation $\varphi:\Psi\rightarrow\{0, 1\}$, aggregation of a binary value of 0 or 1, to object representing the peer's reputation as not trusted or believable. The algorithm may be different in different scenarios, $\varphi(\Psi)=1$ only num.F=0 in the pessimistic algorithm, while $\varphi(\Psi)=1$ if num.S-num.F$\geqq 0$, else$\varphi(\Psi)=0$, in an optimistic method.

Similarly, the feedback message stores set of triples θ={peer.ID, num.A, num.D}. peer.ID demonstrates the source owner's ID, num.A represents the number of accepted times and num.D is for the number of rejected times.

Usually, a threshold $T\in(0,1)$ is set which can be adjusted dynamically in different situations.

$$\Phi(\theta)=\begin{cases}1\; if\; num.A/(num.A+num.D)\geq k\\ 0\; else\end{cases}$$

In Select peer & interaction phase, the Peer R aggregates and ranks resource provider B_ms' reputation, according to received feedback and feedback peers' reputation. This process completely relies on many different technologies. A simple way is to multiply the two factors, that is, with feedback peers' message and reputation. Although this calculation is simple, the result is not accurate. Moreover, in the credibility of simple judgment mechanism for 0 or 1, most information carried by the feedback message is not fully displayed in the process.

3. THE THEORY OF SET PAIR ANALYSIS

In 1989, a Chinese scholar named Zhao, proposed the theory of Set Pair Analysis (SPA). In an uncertainty system, two relative sets are constructed and according to the identity, discrepancy and contrariness, the connection degree of the set pair can be established. SPA theory considering both certainties and uncertainties as an integrated system depicts the certainty and uncertainty systematically.

3.1. The Definition of SPA

Defintion1: Suppose two sets X and Y, $X=\{x\mid\forall x\in X, X\neq\Phi\}$, $Y=\{y\mid\forall y\in Y, Y\neq\Phi\}$

Put them together to form a set pair $H(X,Y)=X\times Y=\{(x,y)\mid\forall x\in X,\forall y\in Y\}$.

$|X|=m$, $|Y|=n$. $|H|=N=mn$. m, n, N, refer to the cardinality of set X,Y,H respectively [10] .

Defintion2: Issue W: The relationship between X and Y

(1) Identity. $x\in X$, and $y\in Y$ are identities on issue w, noted as $xf^{+}y$. $Hf^{+}(X,Y)=\{(x,y)\mid\forall x\in X\;\&\;\forall y\in Y, xf^{+}y\}$ is the identity set of ordered pairs of set X and set Y on issue w. It is obvious that $Hf^{+}(X,Y)\subset H(X,Y)$. Suppose $|Hf^{+}|=S$ is the cardinality of set $Hf^{+}(X,Y)$, and then $a=|Hf^{+}|/|H|=S/N$ is defined as the Identity degree of set X and Y on issue W.

(2) Discrepancy. $x\in X$, and $y\in Y$ are the discrepancies on issue w, noted as xfy . $Hf\;(X,Y)=\{(x,y)\mid\forall x\in X\;\&\;\forall y\;\in Y, xf\;y\}$ is the discrepancy set of ordered pairs of set X and set Y on issue w. Similarly, $Hf(X,Y)\subset H(X,Y)$. Suppose $|Hf\;|=F$ is the cardinality of set $Hf(X,Y)$, and then $b=|Hf\;|/|H|=F/N$ is defined as the discrepancy degree of set X and Y on issue W.

(3) On the contrary, $x\in X$, and $y\in Y$ are opposites on issue w, noted as $xf^{-}y$.

Similarly, $Hf^{-}(X,Y)=\{(x,y)\mid\forall x\in X\;\&\;\forall y\in Y, xf^{-}y\}$ is the contrary set of ordered pairs of set X and set Y on issue w, $Hf^{-}(X,Y)\subset H(X,Y)$. Suppose $|Hf^{-}|=P$ is the cardinality of set $Hf^-(X,Y)$, and then $c = |Hf^-|/|H| = P/N$ is defined as the identity degree of set X and Y on issue W.

$$u(X,Y) = S/N + (F/N)i + (P/N)j \quad (1)$$

is proposed to indicate the connection degree between set X and Y on issue W. In the formula, the term i is the uncertainty coefficient of discrepancy, the term j is the uncertainty coefficient of contradictory, in brief

$$u(X,Y) = a + bi + cj \quad (2)$$

$a,b,c \in [0,1]$ all are real numbers. The IDC (identity, discrepancy, contrary) degree is defined as below,

$$H(X,Y) = Hf^+(X,Y) + Hf(X,Y) + Hf^-(X,Y) \quad (3)$$

Satisfy the normalization condition, so $a+b+c=1$.

3.2. Priority and Theorem

Theorem 1: $u_1 = a_1 + b_1 i + c_1 j, u_2 = a_2 + b_2 i + c_2 j$ and $u_1, u_2 \in U, a_1, a_2, b_1, b_2, c_1, c_2 \in [0,1]$

$$\begin{cases} u_1 = u_2, \text{ while } a_1 = a_2, b_1 = b_2 \\ u_1 \le u_2, \text{ while } a_1 \le a_2, b_1 \ge b_2 \\ u_1 < u_2, \text{ while } a_1 < a_2, b_1 > b_2 \end{cases} \quad (4)$$

Theorem 2:

$$u_1 = a_1 + b_1 i + c_1 j, u_2 = a_2 + b_2 i + c_2 j, ..., uN = aN + bNj + cNj, an, bn, cn \in [0,1],$$

$$a_n + b_n + c_n = 1 (n = 1,2,...,N)$$

(1) Addition rule:

$$u_1 + u_2 + ... + u_N = \sum_{n=1}^{N} u_n = \sum_{n=1}^{N} a_n / N + \sum_{n=1}^{N} b_n i / N + \sum_{n=1}^{N} c_n j / N \quad (5)$$

(2) Multiplication rule:

$$i \times i = i^2 = i, i \times j = j \times i = i, j \times j = j^2 = j$$

$$u_1 \times u_2 = (a_1 + b_1 i + c_1 j)(a_2 + b_2 i + c_2 j) = a_1 a_2 + (a_1 b_2 + a_2 b_1 + b_1 b_2 + b_1 c_2 + b_2 c_1) i + (a_1 c_2 + a_2 c_1 + c_1 c_2) j \quad (6)$$

4. REPUTATION BASED SPA

Due to the importance of different indicators, the average connection degree of the set pair is the arithmetic mean of the very indicator not suitable in the previous studies. By introducing information entropy theory to determine the weights of evaluation indexes, some improvements are made to the original SPA method, which surely provides a new way of thinking and new methods for the evaluation of peer's reputation value.

4.1. The Expression of Reputation Value with Uncertainty

A peer reputation not only expresses its probability of good behavior, but also its probability of uncertain behavior and bad behavior. Suppose,

$P(G)$: The Probability of a peer exhibiting good behavior in transaction

$p(U)$: The Probability of a peer exhibiting uncertain behavior in transaction

$p(B)$: The Probability of a peer showing bad behavior in transaction

$P(G), p(U), p(B) \in [0,1], P(G) + p(U) + p(B) = 1$,

A peer's reputation value is defined by:

$$p(s) = P(G) + p(U)i + p(B)j \quad (7)$$

4.2. The Aggregation of Reputation Value

In this phrase, a peer's reputation value is calculated by the aggregate of received feedback, which is combined with the respondent's reputation value.

Example 1: In one progress, a peer R receives feedback form of peer A for source owner B. A's reputation value is $P(A) = 0.7 + 0.2i + o.1j$, and the feedback value is

$$P_{ab}(A) = 0.5 + 0.3i + o.2j$$

The source owner B's reputation value is calculated by peer R

$$P_a(B) = P(A) \times P_{ab}(A) = (0.7 + 0.2i + 0.1j)(0.5 + 0.3i + o.2j) = 0.35 + 0.44i + o.21j$$

4.3. Scoring and Ranking

Supposing peer R launches a request for reputation value of the resource provider B1 and B2. For each resource provider, if only one feedback is received in the network, then in accordance with the above algorithm, R easily polymerizes each resource provider's credibility and makes a decision whether or not to transact with them in the future. Actually, for each resource provider, R receives a plurality of feedbacks and computes a reputation for each peer, and again these reputations' aggregation is considered as the final reputation.

Example 2: This is 20 peers $a_k, c_k (k = 0,1,2,...,9)$ feedback to the request for B1, B2 from peer R; the combining of ak, ck reputation and feedback value for B1, B2 has been finished respectively as shown in Table **1** and **2**.

R gets the final reputation value of B1, through the following calculation

Table 1. B1 reputation value hold by $a_k (k = 0,1,2,...,9)$.

Pak(B1)	Identity	Discrepancy (i)	Contrary (j)
a0	0.4	0.4	0.2
a1	0.7	0.1	0.2
a2	0.5	0.3	0.2
a3	0.4	0.3	0.3
a4	0.6	0.1	0.3
a5	0.5	0.5	0
a6	0.8	0.1	0.1
a7	0.7	0.2	0.1
a8	0.5	0.4	0.1
a9	0.7	0.3	0

Table 2. B2 reputation value hold by $c_k (k = 0,1,2,...,9)$.

Pck (B2)	Identity	Discrepancy (i)	Contrary (j)
c0	0.4	0.4	0.2
c1	0.6	0.2	0.2
c2	0.7	0.2	0.1
c3	0.8	0.1	0.1
c4	0.9	0.1	0
c5	0.5	0.1	0.4
c6	0.7	0.1	0.2
c7	0.5	0.2	0.3
c8	0.5	0.3	0.2
c9	0.4	0.3	0.3

$$
\begin{aligned}
p(B_1) &= \sum_{k=0}^{9} p_{ak}(B_1)/10 \\
&= (0.4+0.7+0.5+0.4+0.6+0.5+0.8+0.7+0.5+0.7)/10 \\
&+(0.4+0.1+0.3+0.3+0.1+0.5+0.1+0.2+0.4+0.3)i/10 \\
&+(0.2+0.2+0.2+0.3+0.3+0.1+0.1+0.1)j/10 \\
&= 0.58+0.27i+0.15j
\end{aligned}
$$

By the same token: $p(B2) = 0.6 + 0.2i + 0.2j$

In this situation, $p(B2) > p(B1)$. If more resource providers exist, their final reputation value can be obtained from the formula above which is easy to sort.

4.4. The Improved Mechanism

4.4.1. First Make Judge, Before Aggregation

In traditional mechanism, after R receives the feedback messages, R goes through the process of aggregation.

SUPPOSE u, v , are real numbers $\in (0,1)$, $pk(Bk) = a + bi + cj, k \in N$

$$
\begin{cases}
\text{pk(Bm) is indentity, if } \max(a,b,c)=a \geq u \\
\text{pk(Bm) is contrary, if } \max(a,b,c)=c \geq v \\
\text{pk(Bm) is Discrepancy, else}
\end{cases}
\tag{8}
$$

u is the experience value of peer's identity expectation, v is the experience value of peer's contrary expectation. In this method, feedback peers are classified according to the standard after R receives a lot of feedback messages. R first makes a decision on which peer should be given identity to carry out the next calculation.

4.4.2. Index Taken into Account

As mentioned above, to finish a specific transaction, after sending a query message, R receives the response message. This message contains some service quality index, such as: respond time, time consumed, price, *etc*. In order to re-

Table 3. Index value hold by $d_k (k = 0,1,2,...,9)$.

Index	Respond Time	Time Consuming	Quality	Price
d0	×	√	--	--
d1	--	√	×	×
d2	--	×	√	√
d3	√	×	√	√
d4	√	--	--	×
d5	√	×	×	--
d6	√	√	×	--
d7	×	--	√	--
d8	--	√	×	×
d9	--	×	√	√

flect the differences between these indicators, the index is taken into account in the process of reputation aggregation.

Example 3: Suppose peer d_k (k=0,1,...9) sends the value for peer B to peer R, which holds the value of each index, identity by "√", discrepancy by "--", contrary by "×", as shown in Table **3**.

The cardinality of set, $|H| = mn = 40$, $|Hf^+| = 15$, $|Hf^-| = 13$, $|Hf| = 12$

$$R(B) = |Hf^+| / |H| + |Hf| i / |H| + |Hf^-| j / |H|$$
$$= 15/40 + 12i/40 + 13j/40 = 0.375 + 0.3i + 0.325j$$

According to R's interest, different weights can be assigned to each index.

5. EXPERIMENTS AND ANALYSIS

5.1. Comparison of Expectation

Example 4: In Table **1** situation, $a_k (k = 0,1,2,...,9)$ feedback to the request for B1 from peer R, for the new algorithm; the mathematical expectation of the identity values:

$$E(X) = 0.58, D(X)$$
$$= \sum_{k=0}^{9} (X - E(X))^2 P(X) = 0.02172$$

While in original algorithm, $u = a + bi + cj$, if $a \geq \mathrm{T}$, T is threshold, u is identity,

T=0.5 is set, so a_k value is {0, 1, 1, 0, 1, 1, 1, 1, 1, 1}

$$E'(X) = 0.8, D'(X) = 0.16$$

$D(X) < D'(X)$ which shows that the new algorithm is better than the original in stability.

5.2. Comparison of Confidence Interval

Finally, the confidence interval is calculated with 95% confidence probability.

$$I(X) = [\overline{X} - \frac{\sigma}{\sqrt{n}} Z_a / 2, \overline{X} + \frac{\sigma}{\sqrt{n}} Z_a / 2]$$
$$= [0.58 - \frac{\sqrt{0.02172}}{\sqrt{10}} 1.96, 0.58 + \frac{\sqrt{0.02172}}{\sqrt{10}} 1.96]$$
$$\approx [0.58 - 0.09, 0.58 + 0.09] = [0.49, 0.67]$$

$$I'(X) = [\overline{X} - \frac{\sigma}{\sqrt{n}} Z_a / 2, \overline{X} + \frac{\sigma}{\sqrt{n}} Z_a / 2]$$
$$= [0.8 - \frac{0.4}{\sqrt{10}} 1.96, 0.58 + \frac{0.4}{\sqrt{10}} 1.96]$$
$$\approx [0.8 - 0.25, 0.8 + 0.25] = [0.55, 1.05]$$

It is obvious that $I(X)$ is more precise than $I'(X)$, which indicates that the new algorithm is more accurate. The new algorithm can solve the problem of cheating to some extent.

CONCLUSION

The uncertain factors of peer are not considered in the traditional reputation aggregation mechanism. In this paper, the mechanism of aggregation and sorting is improved which can accurately reflect the peers' reliability, easy quantitative calculation and comparison. Empirical results show that proposed approach is scalable, accurate, robust and fault-tolerant.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This Project is supported by the Innovation project of cultivating graduate of Jiangsu Province (Project Number: CXZZ11_0399).

REFERENCES

[1] F. Cornelli, E. Damiani, S. de Captiani di, S. Paraboschi, and P. Samarati, "Choosing reputable servents in a P2P network", In: *Proceedings of the 11th International Conference on World Wide Web, WWW '02*, Honolulu, HI, USA, 2002, pp. 376-386.

[2] K. Aberer, and Z. Despotovic, "Managing trust in a peer-to-peer information system", In: *Proceedings of the 10th International ACM Conference on Information and Knowledge Management,* New York, 2001, pp. 310-317.

[3] R. Zhou, and K. Hwang, "Power Trust: a robust and scalable reputation system for trusted p2p computing", *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 4, pp. 460-473, 2007.

[4] S. Kamvar, M. Schlosser, and H. Garcia-Molina, "The eigentrust algorithm for reputation management in P2P networks", In: *ACM WWW'03,* Budapest, Hungary, 2003.

[5] L. Xiong, and L. Liu, "Peer Trust: supporting reputation-based Trust for Peer-to-Peer electronic communities", *IEEE Transactions on Knowledge and Data Engineering*, vol.16, no.7, pp. 843-857, 2004.

[6] A.Datta, M. Hauswirth, K. Aberer, "Beyond "web of trust": Enabling P2P E-commerce", In: *Proceedings of the IEEE Conference on Electronic Commerce (CEC'03)*, Newport Beach, California, USA, 2003.

[7] A. Singh, and L. Liu, "Trust Me: anonymous management of trust relationships in decentralized P2P Systems", In: *IEEE International Conference on Peer-to-Peer Computing*, 2003.

[8] R. Zhou, and K. Hwang, "Gossip-based reputation aggregation for unstructured peer-to-peer networks", In: *21st International Parallel and Distributed Processing Symposium, IPDPS 2007*, Long Beach, CA, US, 2007.

[9] S. Marti, and H. Garcia Molina, "Taxonomy of Trust: categorizing P2P reputation systems", *Computer Networks*, vol. 50, no. 4, pp. 472-484, 2006.

[10] P. Zhang, and G.Y. Wang, "New theory of set pair", *Journal of Harbin University of Civil Engineering and Architecture*, vol. 33, no. 3, pp. 1-5, 2000. (in Chinese).

Received: June 10, 2015 Revised: July 29, 2015 Accepted: August 15, 2015