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

He Chaokai¹,* and Wu Meng²

¹School of Computer science, Nanjing University of Posts & Telecommunications, Nanjing, Jiangsu, 210003, P.R. China; ²College 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 scores through collecting locally generated peer feedbacks and aggregation. In traditional reputation system, a peer will rate the other by a score, after finished a transaction, no more detail of transaction are considered in the process. In order to solve this problem, we propose a novel reputation aggregation scheme, 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 peers is both the providers of resources and consumers, 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 the characteristics of openness, anonymity, and dynamic nature of peer activities, P2P networks is very vulnerable and easily abuses by selfish and malicious peers [1]. Some peers perform malicious behaviors including providing false services, spreading malicious viruses, and so forth.

In order to encourage peers to participate in the network, combat malicious peers, reputation system take an important role through distinguishing different peers according to their historical behavior [2]. It is obviously that peers which provide reliable services or acted as declared have higher reputation value. With an efficient reputation system, peers don’t 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 great demand to identify trustworthy peers. With the evolution and acceptance of these P2P services, p2p network putting more demand on the efficiency and accuracy of the online reputations system.

In general, after a transaction completed, the participating peer will rate 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 interacted. Before a transaction, a peer will launch a request, by accessing the global reputation scores of the respondents, peers are able to choose the respondent with high reputation value to finish the transaction.

The field of P2P reputation systems has also commanded increasing attention. Reputation aggregation is the most important issue involved in this process, whose function 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 most popular models.

Take 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. five issues that are important to address in P2P reputation system are proposed by The EigenTrust algorithm, which present a distributed and secure method to compute peers’ global reputation values, based on Power iteration. Based on five factors includes feedback in terms of amount of satisfaction, numbers of transaction, credibility of feedback, transaction context factor, 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 trust 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 refers to the collecting infor-
mation on the past transactional behavior of each peer which is also the basis of the reputation system. 2) Aggregating, according to the information collected, reputation system scores and ranks the peers. 3) Punished and reward. Reputation system will take action against malicious peers and rewarding contributors. Each component requires separate system mechanisms as shown in Fig. (1) [9].

![Image](image_url)

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

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

According to the above information, peer need 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 $\geq 0$, else $\varphi(\Psi) = 0$, in an optimistic method.

Similarly, the feedback message store set of triples $\Theta = \{\text{peer.ID}, \text{num.A}, \text{num.D}\}$, peer.ID demonstrate the source owner’s ID, num.A for the number of accepted times, num.D for the number of rejected times.

Usually, set a threshold $T \in (0, 1)$ which can adjust dynamically in different situation.

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

In Select peer & interaction phase, the Peer R will aggregate and rank resource provider’s reputation, according to received feedback and feedback peers’ reputation. This process complete rely on many different technologies. A simple way is to multiplication the two factors, that is, with feedback peers’ message and reputation. Although this calculation is simple, but the result is not accurate. Moreover, in the credibility of simple judge mechanism for 0 or 1, most information carried by the feedback message does 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 contrary, the connection degree of the set pair can be established. SPA theory considering both certainties and uncertainties as an integrated system and depicting the certainty and uncertainty systematically.

3.1. The Definition of SPA

Definition1: 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 $H(X,Y) = X \times Y = \{(x, y) \mid \forall x \in X, \forall y \in Y\}$.

\[ |X| = m, \quad |Y| = n, \quad |H| = N = mn \quad m, n, N, \text{refer to the cardinality of set } X, Y, H \text{ respectively} \ [10]. \]

Definition2: Issue W: The relationship between X and Y

(1) Identity. $x \in X$, and $y \in Y$ are identity on issue w, note as $xf^\gamma^\gamma$. $Hf^\gamma(X,Y) = \{(x,y) \mid \forall x \in X \& \forall y \in Y, xf^\gamma y\}$ is the identity set of ordered pairs of set X and set Y on issue w. It is obviously that $Hf^\gamma(X,Y) \subseteq H(X,Y)$.

Suppose $|Hf^\gamma| = S$ is the cardinality of set $Hf^\gamma(X,Y)$, and then $a = |Hf^\gamma| / |H| = S / N$ is defined to the Identity degree of set X and Y on issue W.

(2) Discrepancy. $x \in X$, and $y \in Y$ are discrepancy on issue w, note as $xf^\gamma^\gamma$. $Hf^\gamma(X,Y) = \{(x,y) \mid \forall x \in X \& \forall y \in Y, xf^\gamma y\}$ is the discrepancy set of ordered pairs of set X and set Y on issue w. similar $Hf(X,Y) \subseteq H(X,Y)$.

Suppose $|Hf^\gamma| = F$ is the cardinality of set $Hf(X,Y)$, and then $b = |Hf^\gamma| / |H| = F / N$ is defined to the discrepancy degree of set X and Y on issue W.

(3) Contrary. $x \in X$, and $y \in Y$ are opposite on issue w, note as $xf^\gamma\gamma$.

Similar, $Hf^\gamma(X,Y) = \{(x,y) \mid \forall x \in X \& \forall y \in Y, xf^\gamma y\}$ is the contrary set of ordered pairs of set X and set Y on issue w. $Hf^\gamma(X,Y) \subseteq H(X,Y)$.

Suppose $|Hf^\gamma| = P$ is the cardinality of set $Hf^\gamma(X,Y)$, and then $c = |Hf^\gamma| / |H| = P / N$ is defined to the Identity degree of set X and Y on issue W.

$$
u(X,Y) = S / N + (F / N)j + (P / N)j$$

(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
\[ u(X, Y) = a + bi + cj \]  
(2)

\[ a, b, c \in [0,1] \] all are real number. We defined the IDC (identity, discrepancy, contrary) degree as below,
\[ H(X, Y) = H^{+}(X, Y) + H^{-}(X, Y) \]  
(3)

Satisfy the normalization condition, so \( a + b + c = 1 \).

3.2. Priority and Theorem

Theorem 1:
\[ u_{1} = a_{1} + b_{1}i + c_{1}j, u_{2} = a_{2} + b_{2}i + c_{2}j \]
\[ u_{1}, u_{2} \in U, a_{1}, a_{2}, b_{1}, b_{2}, c_{1}, c_{2} \in [0,1] \]
\[ \begin{cases} 
u_{1} = u_{2}, \text{ while } a_{1} = a_{2}, b_{1} = b_{2} \\
u_{1} \geq u_{2}, \text{ while } a_{1} \leq a_{2}, b_{1} \geq b_{2} \\
u_{1} < u_{2}, \text{ while } a_{1} < a_{2}, b_{1} > b_{2} 
\end{cases} \]  
(4)

Theorem 2:
\[ u_{1} = a_{1} + b_{1}i + c_{1}j, u_{2} = a_{2} + b_{2}i + c_{2}j, ..., u_{N} = a_{N} + b_{N}i + c_{N}j \]
\[ a_{n} + b_{n} + c_{n} = 1 (n = 1, 2, ..., N) \]

(1) Addition rule:
\[ u_{1} + u_{2} + ... + u_{N} = \sum_{n=1}^{N} u_{n} \]  
(5)

(2) Multiplication rule:
\[ i \times i = i^2 = i, i \times j = j \times i = ij = j \]
\[ u_{1} \times u_{2} = (a_{1} + b_{1}i + c_{1}j)(a_{2} + b_{2}i + c_{2}j) \]
\[ = a_{1}a_{2} + (a_{1}b_{2} + a_{2}b_{1} + b_{1}b_{2} + b_{1}c_{2} + b_{2}c_{1})i + (a_{1}c_{2} + a_{2}c_{1} + c_{1}c_{2})j \]  
(6)

4. REPUTATION BASED SPA

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

4.1. The Express of Reputation Value with Uncertainty

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

\[ P(G) \] : The Probabilistic of a peer act good behavior in transaction
\[ p(U) \] : The Probabilistic of a peer act uncertain behavior in transaction
\[ p(B) \] : The Probabilistic of a peer act bad behavior in transaction

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

A peers’ reputation value is defined:
\[ p(s) = P(G) + p(U)i + p(B)j \]  
(7)

4.2. The Aggregation of Reputation Value

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

Example 1: In one progress a peer R received feedback form peer A for source owner B. A own reputation value is
\[ P_{ab}(A) = 0.5 + 0.3i + 0.2j \]

The source owner B’ reputation value is calculate by peer R
\[ P_{r}(B) = P(A) \times P_{ab}(A) \]
\[ = (0.7 + 0.2i + 0.1j)(0.5 + 0.3i + 0.2j) \]
\[ = 0.35 + 0.44i + 0.21j \]

4.3. Scoring and Ranking

Suppose peer R launched a request for reputation value of the resource provider B1, B2. For each resource provider, if only received one feedback in the network, then in accordance with the above algorithm, R easily polymerized each resource provider credibility, and make a decision whether or not to transact with them in the future. Actually, for each resource provider, R received a plurality of feedbacks, the need for each peer computes a reputation, and again these reputation aggregation as a final reputation.

Example 2: This is 20 peers feedback to the request for B1, B2 from peer R, the combine of reputation and feedback value for B1, B2 has been finished respectively. As follow Table 1 and 2.

R get the final reputation value of B1, through the following calculation
\[ p(B_1) = \frac{1}{10} \sum_{k=0}^{9} p_{ak}(B_1) \]
\[ = (0.4 + 0.7 + 0.5 + 0.4 + 0.6 + 0.5 + 0.8 + 0.7 + 0.5 + 0.7)/10 \]
\[ + (0.4 + 0.1 + 0.3 + 0.3 + 0.1 + 0.5 + 0.1 + 0.2 + 0.4 + 0.3)/10 \]
\[ + (0.2 + 0.2 + 0.2 + 0.3 + 0.3 + 0.3 + 0.1 + 0.1 + 0.1)/10 \]
\[ = 0.58 + 0.27i + 0.15j \]
By the same token:  \( p(B2) = 0.6 + 0.2i + 0.2j \)

In this situation,  \( p(B2) > p(B1) \). If exists more resource provider, their final reputation value can be get from the formula above, easy to sort.

4.4. The Improved Mechanism

4.4.1. First Make Judge, Before Aggregation

In traditional mechanism, after R received the feedback messages, R would going the process of aggregation.

\[
\text{SUPPOSE } u, v \text{ are real numbers } \in (0,1), \quad pk(Bk) = a + bi + cj, k \in N
\]

\[
\begin{align*}
\text{pk}(Bm) \text{ is indentity, if } \max(a,b,c) &= a \geq u \\
\text{pk}(Bm) \text{ is contrary, if } \max(a,b,c) &= c \geq v \\
\text{pk}(Bm) \text{ is Discrepancy, else}
\end{align*}
\]

4.4.2. Index Take into Account

As we mentioned above, to finish a specific transaction, after sending a query message, R received the response message. These message contains some service quality index, such as: respond time, time consuming, price, etc. In order to reflect the differences between these indicators, those index is take into account in the process of reputation aggregation.

**Example 3:** Suppose peer \( d_k (k=0,1,\ldots,9) \) send value for peer B to peer R, which hole the value of each index, identity by “\( \text{\checkmark} \)”, discrepancy by “\( \text{\checkmark}\text{\textbf{\checkmark}} \)”, contrary by “\( \times \)”, show as Table 3.
The cardinality of set, \( |H| = mn = 40 \), \( |Hf^-| = 15 \), \( |Hf^+| = 13 \), \( |Hf| = 12 \)

\[
R(B) = |Hf^+| / |H| + |Hf^-| / |H| + |Hf^+| / |H| \times |Hf^-| / |H|
\]

\[
= 15 / 40 + 12/40 + 13 \times 40 = 0.375 + 0.3i + 0.325j
\]

According to R's interest, different weight 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:

\[
E(X) = 0.58, D(X)
\]

\[
= \sum_{i=0}^{2} (X - E(X)) P(X) = 0.02172
\]

While in original algorithm, \( u = a + bi + cj \), if \( a \geq T, T \) is threshold, \( u \) is identity,

We set \( T=0.5 \), so \( a_k \) value is \{0, 1, 1, 0, 1, 1, 1, 1, 1, 1\}

\[
E'(X) = 0.8, D'(X) = 0.16
\]

\[
D(X) < D'(X) , \text{which shows that the new algorithm is better than the original in stability.}
\]

5.2. Comparison of Confidence Interval

Finally, Calculate the confidence interval with 95% confidence probability.

\[
I(X) = [\bar{X} - \frac{\sigma}{\sqrt{n}} Z_x / 2, \bar{X} + \frac{\sigma}{\sqrt{n}} Z_x / 2]
\]

\[
= [0.58 - \frac{0.02172}{\sqrt{10}} 1.96, 0.58 + \frac{0.02172}{\sqrt{10}} 1.96]
\]

\[
= [0.58 - 0.09, 0.58 + 0.09] = [0.49, 0.67]
\]

\[
I'(X) = [\bar{X} - \frac{0.4}{\sqrt{10}} 1.96, 0.58 + \frac{0.4}{\sqrt{10}} 1.96]
\]

\[
= [0.8 - 0.25, 0.8 + 0.25] = [0.55, 1.05]
\]

It is obviously that, \( I(X) \) is more precise than \( I'(X) \), which indicate the new algorithm is more accurate. The new algorithm can solve the problem of cheating in some extent.

CONCLUSION

The uncertain factors of peer does not considered in the traditional reputation aggregation mechanism, In this paper, we improve the mechanism of aggregation and sorting, which can accurately reflect the peers reliability, easy quantitative calculation and Comparison. Empirical results shows 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


