A Reputation System Model Based on Indication to Combat Pollution in P2P Networks

Fabíola Bento Soares
College of Computing
Federal University of Uberlandia - UFU
Uberlândia/MG, Brazil
e-mail: fabiolasb@gmail.com

Luís Fernando Faina
College of Computing
Federal University of Uberlandia - UFU
Uberlândia/MG, Brazil
e-mail: faina@facom.ufu.br

Newarney Torrezão da Costa
College of Computing
Federal University of Uberlandia - UFU
Uberlândia/MG, Brazil
e-mail: newarney@gmail.com

Jamil Salem Barbar
College of Computing
Federal University of Uberlandia - UFU
Uberlândia/MG, Brazil
e-mail: jamil@facom.ufu.br

***Abstract*— P2P Networks are compound by nodes, servers and suppliers of services or resources. This kind of system allows us as part of it to supply and ask for resources in easy manner as well as for fake or corrupted content resources. Some mechanisms based on resources or based on nodes, named reputation systems, are developed to decreasing the pollution in P2P Network. This article presents a reputation system model based on nodes and also means to defeat the pollution issue by expelling malicious nodes, underprivileging selfish nodes and improving the honest ones. Simulations prove the effectiveness of the model in question.**

Keywords-reputation; poisoning; pollution; peer-to-peer networks

I. INTRODUCTION

A network architecture Peer-to-Peer (P2P) is made of elements that may perform as clients and suppliers of resources [1]. Those networks are spread systems, which have interconnected nodes with self-organization capability, aimed at sharing diverse resources such as, music, video, document, among others. Another skill to be stand out is their capability to adapt at the same time they keep the connectivity with acceptable performance, without mediators or support of a central control office [2].

They also make feasible a great inlet and outlet of members with most diverse intentions, therefore, there are no means to avoid poisoning, that is, the availability of resources with corrupted or useless content by P2P network. Due to this poisoning occurs the pollution, which configures the network breakdown. Such mainly happens for the action of bad-intentioned users that poison a specific resource. This poisoning of resources may take place in several manners such as adding it a invalid content, supplying it with false information able to corrupt files, also by changing frequencies of music files, linking Trojans, among other techniques.

Resources, whether corrupted or not, have the same set of information, thus generating an issue, for it becomes hard to find those not poisoned [3]. Such issue may be detracted by a system of reputation that help the requestor nodes to acquire resources with valid content, besides avoiding dishonest nodes in the network [4]. In this context, there are nodes or selfish users which never share resources or narrow its permission to access, along with generous users that always release valid resources and cooperate for the right maintenance and good working of P2P network by supplying reliable information.

In this perspective, great issues in reputation systems are: discriminate valid resources from invalid ones in distributed and decentralized systems in reason of the dynamic behavior it presents, stand out selfish users to repress them for they don't add resources to the network. This work proposes to minimize these problems by reducing the pollution in P2P networks by means of a methodology which benefits generous users to the detriment of bad-behaved users, whether malicious or selfish.

Therefore, instead of consulting directly the node reputation which hosts the recourse, a consulting to partner nodes is made. These nodes are addressed as supernodes and have already used the recourse requested, or they know the node reputation that hosts this recourse. Thus, the model privileges the good reputation nodes with the information shared with partner nodes. Therefore, the quality of the content offered by network evolves according to the distributed and regular policy. Accordingly, with gradual use of the net, generous users broaden their part in the system due to its quality of resources, giving priority to the proposition of groupware model. Despite the existence of selfish and bad-intentioned users, their performance in the net shall be reduced as partner nodes narrow their reputation.

In order to develop the proposition, this article is structured as follows: the Introduction previously presented; Section 2 approaches the works related; Section 3 describes the model in question; Section 4 presents the simulation to

 ISBN: 978-1-61208-192-2

validate the hypothesis, and eventually, Section 5 brings final considerations.

II. RELATED WORKS

The performance of systems of reputation is based on two main concepts, which are, confidence and reputation. Confidence is about leaving the analysis whether something is a fact or not by delivering this study to the source where this information came from, and simply take this into consideration. For an individual to be considered trustful, it is necessary that it has positive, honest and cooperative attitudes in relation to the entities dependent on it. On the other hand, reliability is one's capacity to be trustful, which means that confidence is a consequence of reliability [5].

Reputation, by one's turn, is what one knows about the character or position of an individual before the judgment of a community. Therefore, reputation reflects the community's vision over an individual, while trustfulness is about a subjective opinion.

The systems of reputation represent a important alternative to help users themselves settling confident relationships through Internet, allowing them to make personal evaluations over individuals performance and identify the reputations estimated before the opinion of a community. Thus, those systems present mechanisms to stand out and manage reliability relations among users [6].

In this work, two systems of reputation were approached that guided the new model proposed: the Credence System [7] and the System Based upon Resources [8]. Both were proposed for P2P environment purely decentralized, in which information of nodes reputation is spread over the net.

A. *Credence system of reputation*

This system allows users to classify the resources obtained concerning its authenticity whether polluted or not. It works based on a protocol of research by voting, used to disseminate the rank of those resources by the system and a correlation scheme of votes that gives more weight to the ones came from pairs that tend to have the same opinion [7].

Before the acquisition of a resource, one computes a correlation among the nodes of the net, that is, applicant and supplier nodes. It is natural that bad-intentioned users lie about their reputation to poison the net. This correlation presents two strategies to protect the statistics of confidence, which is locally stored. In one of them only locally computed correlations are changed, that is, the client may apply for the auditing of the correlation choosing one of the nodes involved on it, keeping its integrity.

In the second strategy, in practice, the local confidence statistics has significant amount of redundant information, densely connected, forming cycles and raising maliciously the reputation before its mates. The auditing might identify such behavior and somehow punish the responsible ones.

In this aspect, the relationship between two nodes is expressed by the correlation of their vote record, checking if the nodes tend to vote in similar manner, which we call positive correlation, in different manner, which is named negative correlation, or if their records of votes are not correlated.

B. *System of reputation based on resource*

In this model, before making the choice of a resource for download, the requestor applies to other nodes the score of the candidate resource, and weighs these resources according to the reliability and information received from partner nodes. When a node requires a resource it consults all the nodes in the net. Once it has the response, the applicant node may choose among the replies and get what it wants. After nodes interaction, the applicant ratify if what was required is in fact what it wanted [8].

This result is not always optimistic. As a common user, it cannot distinguish between the authentic and polluted resources unless it's possible to verify the content after getting the resource, or make a remote evaluation, which is practically impossible. In this system of reputation, each node validates the authenticity of the resource it gets and records its result. Thus, the requestor either receives a authentic resource or a polluted one.

Such mechanism aims at restraining the nodes from diffusing polluted resources in the following situations: a) damage by some kinds of bad behavior, as the sharing of invalid resources; b) false information given to other nodes in the net, thus sending misleading content over the resources; c) collusion of nodes that give right opinions about some resources by pretending to be an honest node in order to gain confidence of another ones.

A distinction is made in relation to the resource and to the nodes concerning the reputation and reliability, in which the reputation of a resource from a node point of view, is used to evaluate the expectancy of a node in relation to this resource. The reliability of a node, from another node point of view, is a subjective expectation which believes that the evaluation of the resource is true.

Each participant node keeps the record of identification of a resource in a set classified as *RS*, and the node identification is classified in a set named *NS*, which is compound by nodes that has been publicly evaluating at least one resource in *RS* set. Each node has a local storage defined by *L* to be saved in a matrix, and the reliability value of other nodes is saved in another matrix *R*.

Each node *P* has a set of resources identifiers RS_P, and a set of node identifiers NS_P. RS_P is compound of resources locally authenticated by *P*. NS_P is compound by nodes which has publicly evaluated at least one of its resources in RS_P, and $RS_P[0]$ is fitted to be *P*.

In addition, each node stores locally the information in which *P* takes part in a matrix L_P. Other reliability values are stored in a vector R_P.

 ISBN: 978-1-61208-192-2

C. Analysis of the reputation systems P2P networks

In both systems analyzed, the polling for a resource brings overhead in P2P net. This occurs because every node which possesses the requested information shall be able to respond to the consult. Also, there isn't a hierarchy model that distributes the nodes according to its reputation. In these systems, it is not considered the validity of the repository. This implies in risks to the security about the broadcast of reputation data in the net. In these models, once the user is considered to be trustful, it shall have the permission to spread the reputation obtained among the network's nodes.

However, for a network with thousands of transactions, in many cases, the access of a same file shall be done by many users at the same time. In case there is a group of malicious users that propagate a information of positive reputation to a set of files with doubtful integrity, this act may compromise the quality of network resources. Therefore, it is interesting that, even with the spread of reputation information, only the average shall be considered. Thus, if 10% of bad-intentioned users pollute the net and forge the reputation of its files, one expects that the evolution itself of the net with interactions among well-behaved, shall be able to judge this poisoning attempt.

The sequence of events to find a resource in both systems is the same. The node requestor, *Node_Requestor*, does the query using keywords, the message *Query(keyword)* is sent to the whole subset of nodes in P2P net, according to the employed protocol. The nodes which don't possess the recourse and are willing to share it, reply with the message *QueryHit(nodeID, resourceID*, see Figure 1.

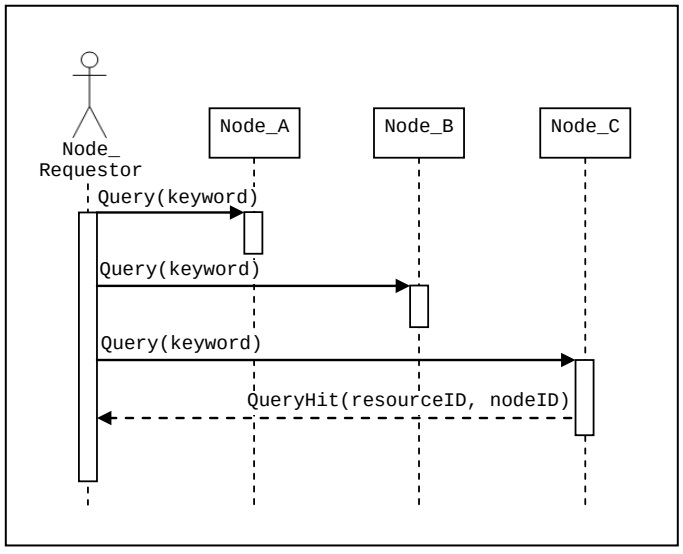

Figure 1. Sequence of events to consulting resource.

After obtaining the responses, proceeds the evaluation of these. The applicant node, repeatedly chooses a recourse by sending a message to request the confidence punctuation for the nodes different from those that have the recourse with the message *RatingQuery(resourceID)*. The expert node responds with a message *RatingHit(nodeID, resourceID, ratingValue)*, see Figure 2.

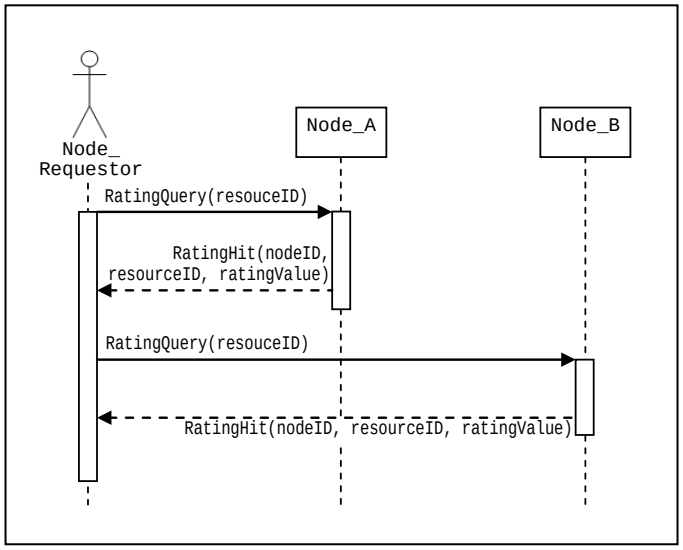

Figure 2. Sequence of events to validate the trust.

At each of the systems, a validation of confidence is done in distinct manner, this correlation is presented next. After the acquisition of the recourse, the nodes which supplied information about the recourses are notified about the achievement of the resource, and the reputation of the recourse origin is updated for higher, when its valid, or otherwise lower.

III. MODEL BASED ON INDICATION

Society represents a model of reputation and confidence evolved, considering that, during the interactions, the behavior of elements belonging to this system is now considered and at each need for interaction, one makes a analysis by which one infers if it is trustful or not to interact with the node. When one needs a specific recourse, the requestor, in its seek of reference data to infer if a behavior is acceptable or not, it consults other elements that may have already acquire such recourse and have good reference of concluded negotiations. This approach differs from current ones, for it proposes to receive a indication of a recourse from a reliable source. Still there is a model of economy that allows the net to stretch, enabling new members to acquire a good reputation in the course of their interactions.

From this information, one applies one's policy of analysis and infers in which node it is possible to acquire such recourse with success. The analogy is established when someone who needs a certain service, a *baby-sitter* for instance, consults somebody who knows a qualified professional that was previously hired and with who one had no issues. This model tends to classify a individual according to its behavior converging to the "true" of the society and not to a pre-established threshold of the system.

The model proposed acts in the same manner of the society, searching recourses by indication of nodes to others they had already interacted. With this information, it's possible to rank the nodes which has similar recourses, and

 ISBN: 978-1-61208-192-2

the requestor policy is uncharged to infer whether one should or should not acquire such recourse, and in addition, classify the transaction and generate the record of interactions, ranking each time more precisely the nodes of the society. The evaluations are made as for the supplier node as to the recourse stored on it, and both has a reputation measured by the system.

Another strategy of the proposed model is to encourage the sharing of recourses. In case it doesn't happen, the malicious of selfish nodes are purged and not allowed to take part in P2P network.

A. Hierachy structure

The nodes of the net are classified in two levels, which are nodes and supernodes. Nodes are the elements that may share and request recourses, and any entrant node in the P2P net fits initially in this category. The supernode, besides being a node, is also able to make indications about the reputation of nodes and recourses, and make public the information of the data repository of its subordinates with the rest of the network. The supernode is a trustful element of the net. Compulsorily, every node is related to a supernode, and this association occurs in the occasion of the inlet of the node in the P2P net. The supernode, chosen to support the inlet node, is the one that possess the least amount of subordinate nodes. A node is associated solely to a supernode [9].

The node is promoted to a supernode when it reaches a pre-established threshold of points. This score is initially pre-defined, but alters dynamically with the evolutions of supernodes, that is, the least score to become a supernode is the average of the current punctuation of a group of supernodes. Besides the scores, one must consider several basic characteristics to promote a node to a supernode, for instance, storage capability, band width, or still, the period of participation in P2P net. After the promotion, it is possible to be lowered to a node, according to its behavior. At each affiliation of a new node to a supernode, this receives a punctuation for publishing the information in P2P network. Every supernode has another contingency supernodes, in case some unavailability, the node may affiliate itself to the contingency supernode [10].

B. Repositories

Locally, each node stores a repository containing information of its shared recourses. In each shared recourse is given a punctuation to the supplier node. Also, it is given a initial punctuation to the recourse. Similarly to the node punctuation, each recourse takes a punctuation at each transaction, which may increase or decrease it. Each local information of each node is replicated in the supernode, and the supernode, with a certain frequency, replicates to the contingency supernodes.

C. Seek for recourses

The seek process for recourses is unique for avoiding the consult to every node of the net, but only to the supernodes, thus reducing a lot the amount of requests sent. When a requestor element, *Node_Requestor*, wishes s certain recourse, it sends to every supernode, including its superior, a requisition *Query(keyWord)*, and every supernode that know the holder of this recourse respond with the identification of it. The supernode, according to the previously mentioned, is a element of reliability that knows either the one who possess a certain recourse or its respective reputation, node and recourse. Thus, the supernode sends such reputations to the requestor node in the message *QueryResponse(resourceID, nodeID, resourceReputation, nodeReputation)*, see Figure 3.

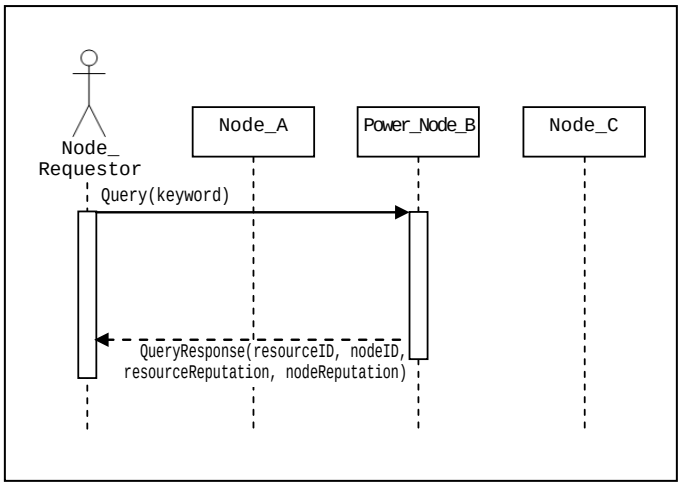

Figure 3. Demand for resources.

Either the score of recourses or the nodes score has a ranking character, since it has positive values, otherwise, it becomes eliminatory. In the course of interactions, in case the node or file zero out its scores, they shall be no longer indicated by the supernode. The nodes which have a score lower or equal to zero may not request for resources as well. This last rule restricts selfish users from staying in the network.

D. Acquisition

With the answers from supernodes, the requestor node chooses the right recourse ordered by the reputation and picks it. The recourses with higher reputation along with the nodes of higher reputation shall probably succeed in the acquisition of the recourse. Each recourse acquired generates a cost in scores for the requestor, independently whether the file is valid or not.

E. Qualification

After the acquisition of the recourse, the node requestor qualifies the transaction, scoring the same way either the supplier node or the recourse acquired. Such interaction is locally stored and sent to the supplier node. In case of conflict of score information, the one that prevails is always the lower, this inhibits bad-intentioned users that lie about its reputation to gain reliability.

 ISBN: 978-1-61208-192-2

IV. SIMULATION

The model proposed assumes that the system may the parameterized in order to attend to the dynamism of the behavior of the P2P networks. To simulate this, values were defined to the score of behavior actions, such as sharing a recourse, consume recourse, among others, according to Table I.

TABLE I. VALUES FOR BEHAVIORAL ACTIONS

Behavioral action	Value
Score to node entrant	100 points for the node.
Scoring for the shared resource	50 points for the resource.
Share resource	1 point for sharing the node.
consuming Resource	-3 Point for the consumer.
Being well qualified to provide resource	10 points for the resource 10 points for the provider node.
Be badly qualified to provide resource	-20 Points for the resource -20 Points for the provider node.
Super-node join node	100 points for super-node.
Score for promotion to the super-node	1000 points

Due to the rule of each inlet node to compulsorily associate to a supernode, it was created a supernode element to start the activities of the P2P networks.

One of the propositions of the model is to permit that new honest nodes get into the net and reach good reputation. After the creation of the supernode to simulate this scenario, one executed 4 steps applied to the values of Table II at each of them.

TABLE II. DISTRIBUTION OF NODES AND RESOURCES APPLIED

Element	Quantity
Nodes generous	35%
Malicious nodes	35%
Nodes Selfish	30%
Total of nodes	2000
Valid files	50%
Invalid files	50%
Total resources shared	56,000
Total interactions made	270,000

The execution considered the selection of requestor nodes and recourses supplied at chance. One of the restrictions imposed was that a node may not take a recourse of itself, for instance, the node itself does not consume its own recourse nor scores its own recourse. The average of reputation of honest nodes raises mainly by the sharing of valid files. On the other hand, the average of reputation of malicious nodes tends to zero, for sharing invalid files. Selfish nodes follow the same trend of the malicious ones, but, with least score, for they only consume recourses, according to Figure 4.

Next, in Figure 5, it is possible to see the evolution of the amount of excluded nodes due to bad behavior and the amount of nodes promoted due to sharing of valid files during interactions. Frequently, nodes with good behavior are promoted once they are never despised, tending to promote every node with this behavior. Malicious or selfish nodes are despised with least proportion to the promoted nodes, still permitting to be promoted in case they become honest nodes.

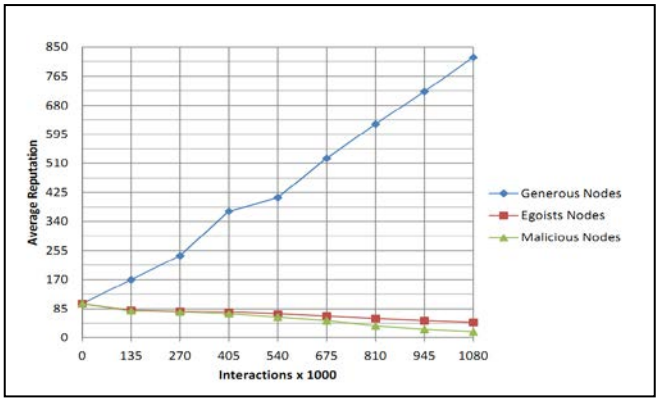

Figure 4. Reputation average x Interactions.

It is important to stand out that threshold of promotion of supernodes are dynamic, and still the scores given to the behavior actions are parameterized, but can be fitted. A node which is promoted to a supernode may be lowered to a node once more in case of a bad behavior, ensuring that the supernodes are trustful elements in the net.

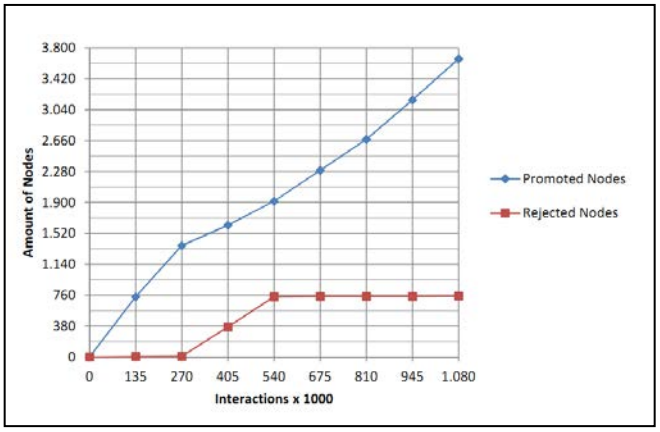

Figure 5. Comparative Interactions x Nodes.

By its turn, evaluations in simulated environment show that the model combats pollution, for attacks mainly the malicious nodes that share invalid files, promoting generous nodes and underprivileging selfish nodes. The benefits that the system promotes were validated.

V. CONCLUSION AND FUTURE WORKS

We presented a model of reputation able to combat the poisoning of recourses and, consequently, the pollution of the P2P network. Still it was showed that, from some interactions with the net, malicious nodes are despised and can no longer share not even consume recourses of the net. Selfish users are also underprivileged and follow the same tendency but with a less steep curve, according to which was proposed in the simulation section.

The hierarchy characteristic applied to the distributed and decentralized net, allows a uniform growth, at the same time that the distribution weighs the supernodes that have less recourses associated. The overhead is drastically mitigated in such a way that, in the worse of cases, it would

 ISBN: 978-1-61208-192-2

reach to be equal to correlated works. The capability to parameterize the system helps finding the best threshold of promotion of supernodes and the purge of malicious and selfish nodes.

REFERENCES

[1] S. Marti, and H. Garcia-Molina, "Taxonomy of trust: Categorizing P2P reputation systems", Computer Networks: The International Journal of Computer and Telecommunications Networking - Management in peer-to-peer systems, vol. 50, Elsevier, August 2005, pp. 472–484, doi:10.1016/j.comnet.2005.07.011.

[2] M. P. Barcellos, and L. P. Gaspary, "Segurança em Redes P2P: Princípios, Tecnologias e Desafios", Proceedings of short courses of the 24th Brazilian Symposium on Computer Networks, pp. 1-50, Brazil, Mai 2006.

[3] C. Costa, "Disseminação de Conteúdo Poluído em redes P2P", Proceedings of the 24th Brazilian Symposium on Computer Networks, pp. 1-13, Brazil, Mai 2006.

[4] J. S. P. Guedes, "Framework para Cálculo de Reputações de Agentes de Software Baseado em Testemunhos", Pontíficia Universidade Católica do Rio de Janeiro, pp. 45-82, Brazil, March 2007.

[5] A. Josang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision", Decision Support Systems, vol. 43, Elsevier, The Netherlands, July 2007, pp. 618-644, doi: 10.1016/j.dss.2005.05.019.

[6] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara, "Reputation Systems: Facilitating Trust in Internet Interactions" Communications of the ACM, Vol. 43, USA, December 2000, pp. 45-48, doi:10.1109/IWQoS.2010.5542754.

[7] K. Walsh, and E. G. Sirer, "Experience with an Object Reputation System for Peer-to-Peer Filesharing", In Proceedings of the 3rd conference on Networked Systems Design & Implementation, Vol. 3, pp. 1-14, USENIX Association Berkeley, USA, 2006.

[8] H. Chen, and G. Chen, "A Resource-based Reputation Rating Mechanism for Peer-to-Peer Networks", Proceedings of the Sixth International Converence on Grid and Cooperative Computing, IEEE Computer Society, USA, August 2007, pp. 535-541, doi:10.1109/GCC.2007.23.

[9] N. Christin, A. S. Weigend, and J. Chuang, "Content Availability, Pollution and Poisoning in File Sharing Peer-to-Peer Networks", Proceedings of the 6th ACM conference on Electronic commerce, ACM, Canada, June 2005, pp. 68-77, doi:10.1145/1064009.1064017.

[10] F. Benevenuto, C. Costa, M. Vasconcelos, V. Almeida, J. Almeida, and M. Mowbray, "Impact of Peer Incentives on the Dissemination of Polluted Content", Proceedings of the ACM Symposium on Applied Computing, France, April 2006, pp. 1875-1879, doi:10.1145/1141277.1141720.

[11] J. Liang, N. Naoumov, and K. W. Ross, "The Index Poisoning Attack in P2P file-sharing Systems", Proceedings of the 25th IEEE International Conference on Computer Communications, Spain, April 2007, pp. 1-12, doi:10.1109/INFOCOM.2006.232.

[12] R. Gupta, and A. K. Somani, "Reputation Management Framework and its use as Currency in Large-scale Peer-toPeer Networks", Proceedings of the Fourth International Conference on Peer-to-Peer Computing, USA, 2004, pp. 124-132, doi:10.1109/P2P.2004.44.

[13] M. Feldman, K. Lai, I. Stoica, and J. Chuang, "Robust incentive Techniques for Peer-to-Peer Networks", Proceedings of the 5th ACM Conference on Electronic Commerce, USA, May 2004, pp. 102-111, doi:10.1145/988772.988788.

[14] S. D. Kanvar, M. T. Schlosser, and H. Garcia-Molina, "The EigenTrust Algorithm for Reputation Management in P2P Networks", Proceedings of the 12th international conference on World Wide Web, USA, 2003, pp. 640-651, doi:10.1145/775152.775242.

[15] A. C. F. Lopes, "Um método para geração de estimativas de reputação mais precisas perante a oscilação de comportamento das entidades avaliadas". Universidade Federal Fluminense, Instituto de Computação, Brazil, July 2006.

[16] C. Costa, "Combatendo a Disseminação de Conteúdo Poluído em redes P2P", Porceedings of the 25th Brazilian Symposium on Computer Networks and Distributed Systems, pp. 1-13, Brazil, Mai 2007

 ISBN: 978-1-61208-192-2