Research Journal of Applied Sciences, Engineering and Technology 7(19): 3995-4001, 2014
DOI:10.19026/rjaset.7.760
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: October 31, 2013 Accepted: November 08, 2013 Published: May 15, 2014

Research Article

Data Sharing in a File Structured QoS Aware Peer-To-Peer System

[1]A. Samydurai and [2]A. Shanmugam
[1]Depertement of Computer Science and Engineering, Dhanalaksmi College of Engineering, Chennai-601301, India
[2]Bannari Amman Institute of Technology, Sathyamangalam-638401, India

Abstract: The Peer-to-Peer system has the potential capacity in building an efficient unified network structure for location based secure data transfer with information sharing and minimizing the failure factor. In this study we define the data transfer with certain futuristic characters that will provide data sharing strength for the users to understand the efficiency with which the data reaches the end node or users. For this purpose we have used specific QoS metric tools like, bandwidth, lookup time, delay, response time and trip time that will help in data sharing. This study concentrates on the search operation on the P2P network for data retrieval. Replication strategy is used in order to increase the probability of successful search. The lookup search employed in this regard uses two type of search namely ring and binary search. The efficiency aspect of data transfer in this study raises the importance of the system compared with the earlier system. The performance analysis will clearly chat out the quality of this system of data sharing and searching. Evaluations have shown a positive sign in the functioning of the proposed system.

Keywords: Content shared model, data sharing, peer-to-peer, QoS, replication, retrieval

INTRODUCTION

Peer-to-Peer network systems are important structures in internet applications. Peer-Peer networks are used to send data from a location to another location providing essential information like bandwidth, lookup time, response time etc. This provides us with the strength of computing large collection of computer networked in a system with preferred cost factor. These peers also help us in receiving among them, without any control, governing the body. The P2P networks have grown into a choice of preference for many researchers to work with, to disclose the search and data retrieval mechanisms occurring in nodes. In P2P networks the data are wide spread and it is stored in the nodes. An efficient searching mechanism has to be employed to search the data from the nodes and extract those data by users, which in turn have to be shared for maintaining the performance and quality in the data transfer (Aleksandra *et al*., 2007; Hai *et al*., 2007; Ea *et al*., 2002).

In this study we are working on a idea where the data are uploaded into the connected P2P networks. The search is performed using lookup operation to find the desired matching data for the user. The user then extracts the data which is to be shared by a sharing mechanism. For efficient and organized data transfers, the peers need to be connected in a structured fashion. Here the peer communication has to be created where all nodes are interconnected and this will ensure data flow through all nodes. The data uploaded provides details of the bandwidth, response time and lookup time of the data transferred to a node. Lookup concept is used in the search operation to locate the data available node in P2P network. Data retrieval efficiencies are high in this aspect. The node failure rate is checked by proximity aware routing (Haiying, 2009). The data retrieval done by users from P2P network should carry a certificate of authentication. This certificate will provide a quality in data transfer and any deviation from proposed standards can be detected. For this purpose we employ machine learning technique and use metrics to read the performances of data transmission and the suitable data can be retrieved. We propose certain figures in this regard to provide a clear understanding of the mechanisms described above. Peer-to-Peer network is built on application layer which is at the top of physical network. But P2P network does not depend on physical network topology. Algorithms are used to organize structured P2P systems. It also uses Distributed Hash Table (DHT) indexing. But no algorithm is used for organizing or optimizing the network connections in unstructured P2P systems (Sergio *et al*., 2007). A P2P system is used to transfer data audio files and any real time data by using ad hoc connections. Actually they are used to connect nodes. The nodes or the system in the P2P system doesn't have

Corresponding Author: A. Samydurai, Department of Computer Science and Engineering, Dhanalakshmi College of Engineering, Anna University, Chennai-601301, Tamil Nadu, India

the relationship of one centralized server and having others as clients. But each and every node in the network acts themselves as client and server simultaneously as needed. P2P is efficient and potential like a client/server model. This creates a virtual connection between the users who are in the network. Cost of implementing a P2P system is very low. It is highly reliable. P2P system can be easily scaled. Multipoint data transfer and communication is efficiently done in P2P systems. The nodes or systems in P2P networks share the data from one end to other end without any difficulty. File sharing and resource sharing is done efficiently and effectively. When resource is transferred with some bandwidth, memory spaces and computation properties, the capacity of the system in turn gets increased because of the increase in system's demand. The P2P networks are robust in nature. The peers need not depend on any centralized server. The P2P networks are much safe and secure while compared to other networking techniques (Gang *et al*., 2008).

The sharing tools like response time (Kalman *et al*., 2009; Tian *et al*., 2009) bandwidth usage and variation in bandwidth usage are used for monitoring and measuring efficiency of P2P system. The sharing of P2P systems is not a complex process which can be done easily without having any changes in satisfying the quality objective requirements. The sharing system of P2P networks connects all the nodes in the system and combines them and establishes new system configurations. Quality sharing can be done by measuring response time, availability of data hop count per lookup etc., various metrics above mentioned are used for sharing P2P system. In a certain P2P systems replication of objects takes place at nodes even though the node has not requested the object. In such cases the search overhead for the object cannot be minimized. We can address these problems by evaluating algorithms using simulation. It is clear that uniformly replicating all objects, even those that are not frequently queried are inefficient (Sergio *et al*., 2007). A more natural policy is to replicate it proportionally to the querying rate which is given by the proportional and uniform replication strategies which yield exactly the same average search size. The average way to allocate the replicas so that the average search size is minimized can be found using square root replication method. The average search size is by obtaining truncated Pareto distribution in which the optimal average search size becomes constant for where, there is a shape parameter. There are two replication strategies that are easily implementable. One is "Owner Replication", where, when a search is successful, the object is stored at the requester node only. The other is "Path Replication", where, when a search succeeds, the object is stored at all nodes along the path from the requester node to the provider node (Sergio *et al*., 2007; Ea *et al*., 2002; Fina and Shenker, 2002).

LITERATURE REVIEW

Monitoring provides current status about the quality metrics of a P2P system such as response time data availability, hop count per lookup, bandwidth consumption, local storage consumption (Kalman *et al*., 2009). Distributed Approximate System Information Service (DASIS) acts as a tool to aggregate information on the state of a P2P system with the minimal depth information obtained from DASIS "Depth Join" algorithm resulted in better balanced P2P system (Keno *et al*., 2004). Structured P2P overlays, which provide a useful substrate for building distribution along with MSP as try technique, can achieve good performance and high dependability in realistic environments with high churn rate (Miguel *et al*., 2004). A hybrid method based on analyzing the basic method iterative key-distribution between higher and node is used for fast and cost effective load-balancing in distributed system that support range queries (Chyouhwa and Kun-Cheng, 2008; Ioannis *et al*., 2009). Moritz *et al*. (2007) discussed about organizing peers with the help of distributed hash tables. In real time systems a very little DHT's were implemented in large scale. Details about past peers up time can be used to calculate remaining uptime. Hai and Liang (2008) discussed about replication of files that helps to improve the efficiency which is used for the replication of the file. Swarm technique increases the effectiveness and efficiency while compared to other methods for file replication. This contains about distributed hash table its development, how it is implemented and simulated on PeerSim for Peer-to-Peer networks (Sergio *et al*., 2007). This proposal is about controlling free riding in Peer-to-Peer systems with the help of distributed and sharing based mechanism which views the Peer-to-Peer network as a distributed and a monitoring structure which emphasis on controlling misbehaviors in peer-to-peer network (Tian *et al*., 2009). In a decentralized unstructured P2P networks resultant for the query is obtained through flooding and random walks. The network traffic can be reduced by the orders of magnitude resolving multiple random walks (Ea *et al*., 2002). In it, is a proposal about the adaptive termination, small granularity of coverage and minimization of message duplication which are the three properties of scalable search algorithm. Various (Agosti *et al*., 2008) P2P models have been defined in the recent years using simulation tools and are evaluated. This proposes a large-scale overlay network simulation through churn modeling. P2PAM uses enhanced node date structures and automates the simulation. Network dynamics such as churn, duplication and search is simplified using scripting language. In the replication strategy used is optimal and it lies in between uniform and proportional. The optimal strategy can be achieved by using simple distributed replication algorithm. The optimal replication policy becomes a square root replication when it is restricted to soluble queries (Fina and

Shenker, 2002). Proposal Alex *et al*. (2002) the explanation for the distribution of degrees in the internet topology, using power law. Similar distributions like number of hopes per message, largest eigen values in the internet graph are also observed. Gives Michalis *et al*. (1999) a novel perspective of the structure of the internet through simple power laws which describe the skewed distributions of graph. Power laws can also be used for the estimation of average neighborhood size and facilitation of design and analysis of protocols. They quantify the highly skewed distributions by single numbers, by power-law exponents. Proposal Noam *et al*. (2007) "scale-free" random networks based on measurements indicating that many large real-world networks have certain scale-free properties. The nodes are the point's chosen uniformly at some random locations. The trade-off is between the geometric considerations that are desirable to connect it to the "central" node in the network. Centrality is measured using the graph distance to the initial node.

MATERIALS AND METHODS

The system architecture of P2P data transmission explains the data uploaded and download into the peers with a graph representation of the peer nodes and how the connection can be established between the nodes that create a structured P2P network we consider a six peer nodes and we are establishing the P2P connection between them. In Fig. 1 we have six nodes N_0, N_1… N_5. Graph G_1 is formed by lacking union of nodes 1 to 5 with its vertex V_1 U V_2, V_2UV_3, V_3 U V_4 … so on.

We assume that node 0 is our source node (say). We have to send data from source node to any one node in the peer say N_2 or to all nodes. To connect the source node with other nodes, the union operation is performed between the source node and graph G_1. In Fig. 2 gives a structured network of Peer-to-Peer nodes with source node connected to every other node which enables data transmission in the network. If many nodes are involved in the P2P network, graph G are formed and are checked for isomorphic characteristic of graph.

Data sharing metrics in P2P network: In P2P network the nodes are interconnected to enable a structured and graph network. The nodes are connected in order to transmit the data between them. When a node sends data to another node, the P2P network, monitoring its quality metrics has to establish the data quality in the network. We consider certain parameters that guide us in monitoring the quality. Source, destination, status, Bandwidth, lookup time, delay, response time and trip time are monitored for each state sent in P2P network. These metrics are studied for analyzing the interrelationship that can be established between each metric parameter we employ the machine learning technique where the performance, bandwidth and storage of the data sent are taken into consideration.

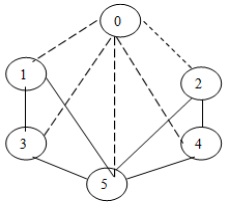

Fig. 1: Source node to be connected with neighboring nodes

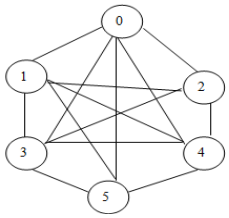

Fig. 2: Connected nodes ready (G1 U N_0) for data transmission (G_1+N_0)

Assumptions:

D → Original data
D_L → Data Loss
C → Computing Value
B → Bandwidth Metric
S → Storage Metric
C_{LB} → Lowest bandwidth value
C_{LS} → Lowest storage value incoming
P_B → Present bandwidth value of data
P_S → Present Storage value of incoming data

We suggest a benchmark criteria of low bandwidth, low storage for optimal performance. To compute bandwidth and storage metric value of 'n' data. This criterion is compared with the incoming data from the source node and the appropriate data is transmitted to neighboring node. This machine learning technique helps us in sharing the suitable data between the nodes. As a result we can bring out the best possible result displaying the quality metric parameters and associated value:

1 if (D_L = = 0)
2 {
3 $C = \sum_{i=1}^{n} c_i(B,S)$
4 $I_1 = C_{LB}, I_2 = C_{LS}$
5 $if(P_B > I_1 \,\&\, P_S > I_2)$
6 D = Less optimal performance;
7 Else if
8 D = optimal performance;
9 Else
10 Retransmit data;
11 }

We compute the bandwidth and storage metric parameter values for 'n' data. We check for the optimal

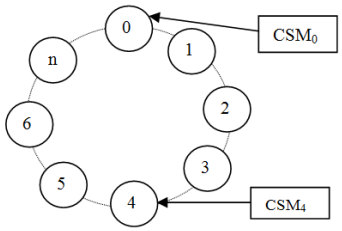

Fig. 3: Ring search topology

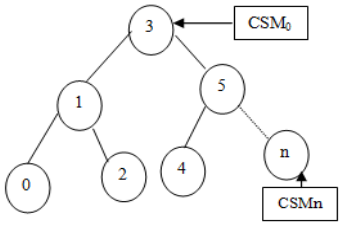

Fig. 4: Binary search topology

performance in each data. If the data loss in the transmission is zero, then we get a optimal performance data. For this purpose, we collect the bandwidth and storage values for n data. If we take the lowest bandwidth and storage parameter value and compare with present value data. If the presented data value satisfies the condition, then the data has optimal performance or else it is requested for retransmission of data.

Topology based search: According to the search topologies in Fig. 3 and 4, each peer in the network has its own CSM_0 (Content Shared Model) which shares the metrics of source peer which are used to locate the data, the user needs and retrieves it efficiently. The search process involves the use of lookup operation where the IP address and port of a peer are determined and is sent to the super peer. The super peer has many child peers. The super peers will request the child peers to search for the particular data in the P2P network. This lookup operation will list the matching peer that contains the data. The user can get the relevant data directly from the child peer through super peers. This type of search is very efficient and reliable in retrieving complete set of data.

RESULTS AND DISCUSSION

When the numbers of nodes in the network increase to a great extent, which is a rare phenomenon, the degree of distribution tends to follow the Poisson distribution, in order to achieve minimum failure. The complexity of ring topology is observed as O (N), whereas the binary search tree time complexity is O (log n) which is considerable less compared to ring search. Therefore binary search is considered as the best search in this case. The probability (Agosti *et al*., 2008) is given as:

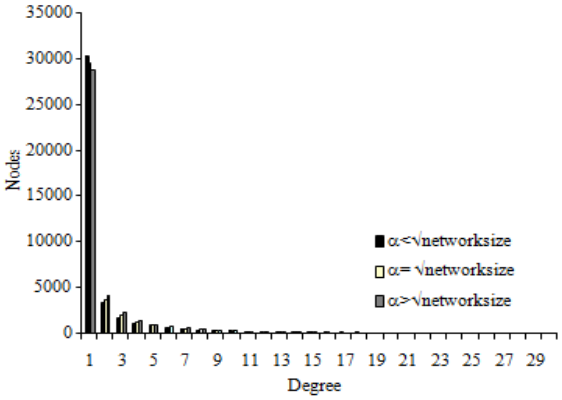

Fig. 5: Topology and degree distributions with three cases of alpha

$$P(k)=\frac{\beta^k e^{-\beta}}{k!} \text{ with } \beta =< k > \quad (1)$$

(i.e.,) average number of nodes connected to one node.

To represent such a scenario we go for random graph model such as Erdos and Renji model, Watts and Strogatz model.

For instance of such a scenario, we take into account a network whose size is 40,000. The various degree possible in that network and the number of nodes having a particular degree are simulated using PeerSim simulator. The simulator varies with α value for a constant size in three different manners (Gian, 2006):

- When $\alpha < \sqrt{networksize}$, the network forms clusters and if α value is very low, the topology tends to be star topology.
- When $\alpha = \sqrt{networksize}$, the network topology is linear.
- When $\alpha > \sqrt{networksize}$, no clusters are formed.

The degree distribution is graphically represented in Fig. 5, "topology and degree distribution with three cases of α". Here we take network size as 40000, α = 175, 200 and 225 for the formally mentioned cases, respectively.

Metrics to satisfy QoS requirements: QoS metrics like bandwidth, lookup time, delay, response time, trip time must be considered with much care in data sharing. When a client searches for a particular data, the probability of success must be the maximum for achieving the required QoS. Also the time client waits for the requested file must be bare minimum. The uploading of the file is based on the bandwidth and the size of the file. These metrics must be satisfied so as to achieve a fulfilling QoS.

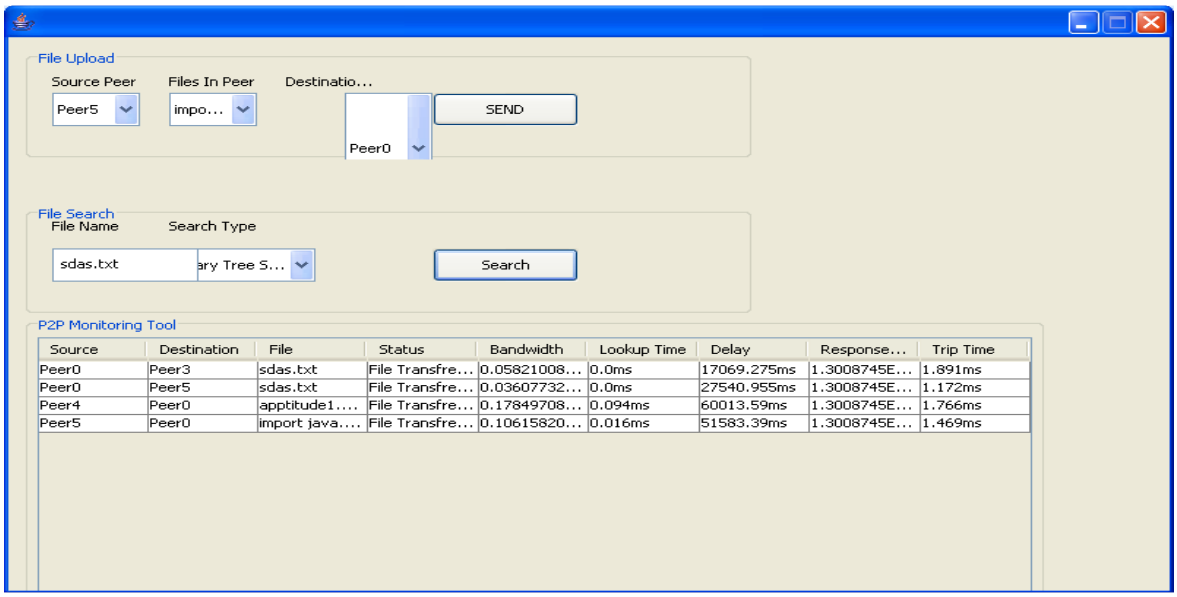

Fig. 6: QoS aware data sharing model

Figure 6 shows the view of real time application of the proposed paper. The Figure displays three component of the paper namely file sharing, file search and Monitoring of QoS in P2P. From the interconnected peers of P2P network we select a peer as the source peer and use the given information about the files contained in the source peer for transferring or data sharing. The distribution peer is selected. In this process the IP address of distribution peer is requested, since different peers have different IP addresses only when the IP address is given the data is sent to the destination. The second component is the file search where the data files of user's interest are searched in the P2P network. In this case the file name and the type of search are given using the search operation, here the request for IP address is to enable search in all the nodes. The user can choose between the ring search and binary search for locating the data. The third component is the monitoring aspect of data file sharing between nodes, when the data are transferred between nodes or peers; the metric tools are displayed that provide information about the data sharing process. The metrics like bandwidth, response time, delay and look up time are calculated for efficiency analysis on data transfer. The performance analysis provided above gives a clear idea on the statistical performance of each node and the search operations on each node in P2P network. In Fig. 6 shown that the chart has those two parameters (search type). Six nodes are considered by taking two search types into consideration. The clearly indicates that the search operations (both binary and ring search) takes a varied time period for search. The Binary search maintains a specific range showing lesser fluctuation. But the ring search shows drastic changes in the time range. Hence we can infer that the larger the node size higher the time taken for search mechanism. But here the binary search shows better results than ring search but still both are efficient search techniques and can be used for location based search of the data.

Replication strategies: In a Peer-to-Peer network, where an object is replicated at some random locations, the searching of that object can be achieved through random walks. An object is said to be replicated when its copy resides in a number of nodes. Replication may be uniform, proportional, square-root, zift-like, path replication, random replication and so on. For updating replications we take into account the node degree, which follows a geometric distribution, whose probability mass function (pmf) is defined as follows (Fina and Shenker, 2002):

$$P(X = ur) = q^{ur-1}p; \{ur = 1, 2, 3, ...\infty\} \quad (2)$$

where,
ur = Number of replications and p + q = 1

The success or failure update of a link between nodes is not taken into account here since that probability is independent of the other links existing in the system. Consider there are n objects in T_n sites. Each object i is randomly replicated at sited ur_i, which are distinct. The query and replication strategy are such that the search continues until a copy is found. The probability for update replication is mathematically given by:

$$Pur_i(k) = \frac{ur_i}{T_n}\left(1 - \frac{ur_i}{T_n}\right)^{k-1} \quad (3)$$

For example we take a P2P network where the total number of sites is six; the probability for update

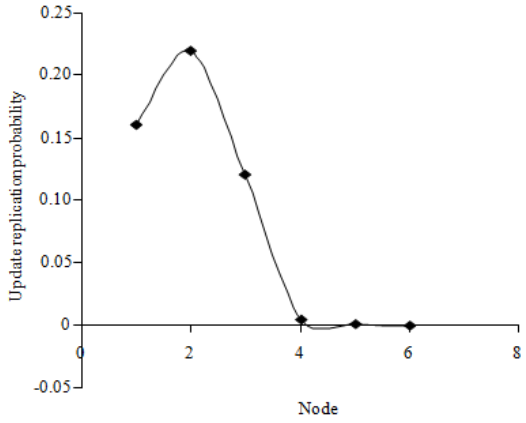

Fig. 7: Updating probability for each node in the P2P network

replication using the above formula is obtained as 0.16, 0.22, 0.12, 0.004, 0.0006 and 0.

A Graph is plotted with nodes against the calculated probability value as shown in the Fig. 7, "Update probability for each node in the P2P network". We observe that the probability value for last node is always zero, since there is no limit on ur_i and hence the search is trivial techniques and can be used for location based search of the data.

CONCLUSION

In this study we have worked on the sharing and optimized search and data sharing in structured Peer-to-Peer network. The work was on the data transfer between the nodes in a structured network and thus calculating the performance efficiency by considering certain metric. The use of metrics like bandwidth, response time, look up time etc., has defined the data transfer with statistical calculation of performance. The search in P2P network has been enhanced by providing through best search techniques for the users to retrieve the user interest data file from the nodes of P2P network. The advantage of this P2P networks is that the nodes are not controlled by any central governing body. This provides a independent data movement between the nodes. The performance analysis done on this mechanism has yielded a positive result from it. The performance efficiency is higher in this aspect and is considered an efficient technique of data retrieval/transmission under the sharing, searching and controlling.

REFERENCES

Agosti, M., M. Amoretti, F. Zanichelli and G. Conte, 2008. P2PPAM: A framework for peer-to-peer architecture modeling based on PeerSim. Proceeding of 1st ICST/ACM International Conference on Simulation Tools and Techniques for Communications, Networks and Systems (SIMUTools 2008), pp: 1-10.

Aleksandra, K., L. Nicolas and S. Ralf, 2007. Globase.KOM-A P2P overlay for fully retrievable location-based search. Proceeding of the 7th IEEE International Conference on Peer-to-Peer Computing, pp: 87-94.

Alex, F., K. Eliasn and H.P. Crististos, 2002. Heuristically optimized trade- offs: A new paradigm for power laws in the internet. Proceedings of the 29th International Colloquium on Automata, Languages and Programming (ICALP '02), pp: 110-122.

Chyouhwa, C. and T. Kun-Cheng, 2008. The server reassignment problem for load balancing in structured P2P systems. IEEE T. Parall. Distr., 19(2): 234-246.

Ea, Q., L. Pei and C. Edith, 2002. Search and replication in unstructured peer-to-peer networks. Proceedings of the 16th International Conference on Supercomputing (ICS'02), pp: 84-95.

Fina, E.C. and S. Shenker, 2002. Replication strategies in unstructured peer-to-peer networks. Proceedings of the 2002 Conference on Applications, Technologies, Architectures and Protocols for Computer Communications (SIGCOMM'02), pp: 177-190.

Gang, C., P.L. Chor and Y. Zhonghua, 2008. Coordinated service provision in peer-to-peer environments. IEEE T. Parall. Distr., 1(4): 433-446.

Gian, P.J., 2006. Peer-Sim HOWTO: Build a Topology for PeerSim1.0. Retrieved form: http://peersim. sourceforge.net. pp: 1-11.

Hai, Z. and F. Liang, 2008. Distributed suffix tree overlay for peer-to-peer search. IEEE T. Knowl. Data En., 20(2): 276-285.

Hai, J., Y. Hong, L. Xiaofei, Y. Sirui and L.Y.J. Wei, 2007. PKTown: A peer-to-peer middleware to support multiplayer online games. Proceeding of IEEE International Conference on Multimedia and Ubiquitous Engineering (MUE'07), pp: 54-59.

Haiying, S., 2009. Efficient and effective file replication in structured P2P file sharing systems. Proceeding of IEEE 9th International Conference on Peer-to-Peer Computing (P2P '09), pp: 159-162.

Ioannis, K., T. Dimitrios and K. Nectarios, 2009. Measuring the cost of online load-balancing in distributed range-queriable systems. Proceeding of 9th IEEE International Conference on Peer-to-Peer Computing (P2P'09), pp: 135-138.

Kalman, G., S. Dominik, R. Julius and K.R.S. Aleksandra, 2009. Monitoring and management of structured peer-to-peer systems. Proceeding of 9th IEEE International Conference on Peer-to-Peer Computing (P2P'09), pp: 311-320.

Keno, A., A. Auedi, G. Michael and W. Roger, 2004. Aggregating information in peer-to-peer systems for improved join and leave. Proceedings of the 4th IEEE International Conference on Peer-to-Peer Computing (P2P'04), pp: 227-234.

Michalis, F., F. Petros and F. Christos, 1999. On power-law relationships of the internet topology. Proceedings of the Conference on Applications, Technologies, Architectures and Protocols for Computer Communication (SIGCOMM '99), pp: 251-262.

Miguel, C., C. Manuel and R. Antony, 2004. Performance and dependability of structured peer-to-peer overlays. Proceeding of the IEEE International Conference on Dependable Systems and Networks (DSN'04), pp: 9-18.

Moritz, S., E. Taoufik and W.B. Ernst, 2007. Analyzing Peer Bebavior in KAD. Eurecom Sophia Antipolis. pp: 1-14. Retreieved form: http://research. microsoft.com/.

Noam, B., B. Bela, B. Christian, C. Jennifer and R. Oliver, 2007. Degree distribution of the FKP network model. J. Theor. Comput. Sci., 379(3): 306-316.

Sergio, B.M., E. Correia and B. Pedro, 2007. Storage and retrieval on P2P networks: A DHT based protocol. Proceeding of 12th IEEE Symposium on Computers and Communications (ISCC, 2007), pp: 623-629.

Tian, J., L. Yang, J. Li and Z. Liu, 2009. A Distributed and Monitoring-based Mechanism for Discouraging Free Riding in P2P Network. Proceeding of Computation World: Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns (COMPUTATIONWORLD '09), pp: 379-384.