Indian Journal of Science and Technology, *Vol 10(38), DOI: 10.17485/ijst/2017/v10i38/98104, October 2017*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

An Efficient p2p Content Retrieval in Guided Distributed Environment using Dynamic Search Algorithm

M. Rajavel* and S. P. Maniraj

Department of CSE, SRM University, Chennai-603203, Tamil Nadu, India;
rajavel.m@vdp.srmuniv.ac.in, maniraj.p@rmp.srmuniv.ac.in

Abstract

Background/Objectives: The Peer-to-Peer (P2P) network connection can build loosely coupled application that covers effective resource sharing. The main aim of this propounded approach is to develop the search and recoveryoperation in the unstructured p2p networks. **Methods/Statistical Analysis**: The network can be classified either as structured or unstructured. Since, the p2p networks which are mainly unstructured can be built efficaciously and with more precision, even in severe conditions over the network topology;so as a result, these networks are far more suitable to the network environment. The random search techniques adopted by these types of networks commonly perform less effective in large size network. The search and retrieve performance can be improved in unstructured p2p networks by utilizing users common Preference pattern, which will be captured inside a probability-theoretic structure is called User Preference Model (UPM). It is developed with the support of Dynamic Query Retrieval Protocol. **Findings:** In our propounded approach, whenever the same file is searched for over three times, the details, present in the routing table will be automatically moving to the User Preference Model Table, hence rendering our propounded searching methodology much faster and efficacious in nature. **Improvements/Applications:** The application of the research is to be used in the small world unstructured peer to peer networks. The search performance and retrieval performance can be improved in this research. **Novelty:** In this specific paper, we have propounded a search protocol as well as a routing table updating protocol so as to develop or improve the searching process through self-managing the p2p network structure into a small business world application.

Keywords: Dynamic Query Retrieval Protocol, Unstructured Peer to Peer Network, User Preference Model

1. Introduction

A lot of search algorithms have been implemented in unstructured p2p networks, with the aim of improving the efficaciousness and performance of the searching and retrieving operation. One of the major challenges in peer to peer networks is the designing of efficacious algorithms for thequicker searching and recovery or retrieval of result from the destination. The two most important searching algorithmsthat wereusedantecedently were Flooding algorithm and Random Walk algorithm. In the case of Flooding, abulk amount of query messagesare generated, hence it becomes uncountable, eventually leading to hiking in the search cost. In the case of Random walk, which is nothing but a conservative search algorithm, where only one node is visitedin the case of each single hop. Hence, by the implementation of random walk algorithm, the search cost was considerably reduced, but it resulted in an increase of time for searching and retrieving the result, as well as multi-hop routing became impossible[1,2]. Hence, as a result, these algorithms were not efficient and were not suitable for the approach, so in order to ensure the efficaciousness of the searching methodology, dynamic search algorithms have been implemented by us in this propounded approach. Dynamic search algorithms are

**Author for correspondence*

implemented to avoid the unnecessary additional over-headsrelated to the searching and retrieval of the result. It will also result in the decrease of time needed for the process in the unstructured p2p networks.

Until and unless, any severe constraints, restrictions or rules have been mentioned in the case of the formation of the network topology, unstructured p2p networks[3,4] can be built very efficaciously, which has already resulted in the involvement of a huge number of live applications usage on the Internet when compared to the structured networks. The Peers in the networks are frequently termed as the blind node, as they are generally ineffectual to find out the opportunity that their nearby peers can promise its resource queries. The vital plan of this research is that the statistical structure over locally that means nearly shared resources and information of peers can be used to lead content retrieval and also improve the resource location detection performance in this Peer to Peer networks. In order to avert from the blind search methodology, three important research issues have been identified in our propounded approach.Peers' Content files are the primary type of resources, which will be considered of prime importance in our propoundedapproach. The initial research difficulties question the reasonableness of creating users' dissimilar preferences. To prevent this difficulty, here we have initiated User Preference Model (UPM) with the help of Dynamic Query Retrieval Protocol[5,6]. By the usage of UPM, the probability value of any peer sharing a particular resource files fs and also sharing of one more resource (file) fd can be pre-estimated. This pre-estimation subsequently gives rise to preference distance between two peers. The second mostissue measures the actual study of users' preferences as incarnated by the User Preference Model. Dynamic Query Retrieval Algorithms can be used for this purpose. Every time, a query is received by a peer for a certain file location, which is not available in its own routing table or User preference Table, then the requested query is forwarded to all its neighbors, having the next topmost probability of sharing their own file.

The last research problem deals with the perceptual experience that the search and retrieval methodology are not enough to accomplish query retrieval operation and high resource exploration. In this specific paper, we have propounded a search protocol as well as a routing table updating protocol so as to develop or improve the searching process through self-managing the p2p network structure into a small business world application. The updation of the routing table should be done, so as tominimize the overall cost required for processing and communication, which are estimated from the received queries in each peer[1,2]. In p2p network, request queries[7] can be fulfilled by any of the peers in the peer network with different probability values.

In this propounded approach, specifically, three protocols of competing strategies have been mentionedin order to update the routing tables by the help of theoretical explanations[8,9]. Our propounded research protocol outplays the rest of protocols based on the result generated, which incorporates Enhanced Clustering Cache Replacement (ECCR) and Least Recently Used (LRU). Under different network conditions, the research has been executed and are reported, which includes network sizes, various size of routing tables, and imprecise UPMs. Our protocols strength is further measured in a dynamic environment, in which the peers join and leave from the p2p network endlessly. These research results authorize that the models and protocols presented here are very effective.

There are numerous modern methodologies available for finding the resource locations in the p2p network environment. Initial approaches to learning like indexing and searching objectives in peer to peer network architecture have been designed[10,11]. Due to the absence of quantify ability and single point of failure issue, late research in the Peer to Peer are firmly pointed to distributed search technology and techniques. There are two fundamental classifications of p2p systems, which have been proposed to affirm distributed search process developments, which are also usually termed as organized or structured networks and unstructured p2p networks[11,12]. Infrastructure technologies with huge practical values will be considered broadly in the structured p2p networks. These features are covered in the form of Distributed Hash Tables[7]. The New files which have to be sent to the pointed peer of the network are related to well-designed and prepared topology rules. The conditions are further used to make users query in the network. The structured Peer to Peer networks can break the most important independent assumption that is broadly accepted in the internet environment, as peers are necessary to hold files which are not preferred by the users mostly. The maintenance of such networks may also lead to increase in cost in highly dynamic environments.

2. Materials and Methods

Peers in the unstructured Peer to Peer networks utilize more autonomy to choose their neighbor's peers and locally shared information or files. An entirely P2P networks in the unstructured pattern such as Gnutella, blind search via flooding mechanisms[13] is usually discovered for finding the resource location and discovery. The peer should be sending the query to all of its neighbors on the network environment in order to fulfill the objective of finding and retrieving a file. Those neighbor peers, i.e., in turn, the peer will be forwarding the query which is requested by the source peer, to all of its neighbor's peer in the network until the query has covered a specific distance. Irrespective of its robustness, simplicity and flooding methods, do not measure the values. In such huge networks, the possibilities of a hit search may be decreased intensely without drastically increasing the flooding distance. With the aim of improving search performance, guided search must be used instead of flooding[13-18]. The main quandary is regarding what all files, in general, are actually qualifiedfor guiding the search process.

The basic perception is that if a peer p2 has some file which is needed by another peer p1 in the same network, then p2 is most certainly to have some other files to be requested by p1in the future. P2 may have the stronger possibilities of having the needed file after the successful retrieval in the first query request. So, in accordance with the previous successful queries, shortcuts from peer p1 to peer p2 are established so as to ensure acceleratingof successive search processes. Similar contemplation and methods have been explored in some other concepts. The intrinsic interest of a particular peer has been dignified within linguistic structure[16]. Few relationship methods, based on these linguistic structures of different formats are again demoralized to find out the two peers or tasks distance in the network. The distance measure factors are generally used for taking numerical values. These values will be an aid in guiding the search process in the upcoming future. No peer should be similar to each other based on the similarity relationship measure. This paper initiates a probability-theoretical structure for capturing users' common preferences.

2.1 P2P Network Architecture

The propounded approach closely conducts a probe on a loosely coupled unstructured p2p network. The architecture of p2p network can be easily understood with the help of Figure 1. Here we are using the letter P to specify a single peer in this p2p network, and P is further used to specify the group of peers available in the network. Data or information file is specified by f. The group of files shared by peer p for every peer p is specified by the variable F_p. To conduct distributed search in the p2p network, every peer p that holds and manages a database itself, having the information of neighbor peers, are connected to the network environment. This list acts as a routing table for the mentioned peer p, which Routing table is denoted by R_p. An upper bound value, Br, present in the peer's routing table, which represents the routing table size, $R_{p,}$ the count of entries represent the size of the routing table R_p. An entry valueE_p of $R_p$is actually an ordered pair consisting of two elements: <f1,p1>. This denotes the link from one peer p to another peer p1 that shares the file f1.

To find (or locate) a data or information file, the peer p user, specified by up, that sends the request query to their network environment. Peer p originates the query which is denoted by q_p and is an ordered set of six basic important elements: <p, f, h_q, TTL, t_s,t_e>. In this elements, the letter p denotes the peer which sends a query q_p, and the file requested by the query $q_p$is denoted by f. The recorded history of searching files is denoted by h_q, which has the details about the list of peers that the query have already processed previously, along with peer p. The TTL has been maintained to achieve a successful query processing. If the time exceeds the level means it will abort or fail the transaction. Within the time limit, the transaction needs to happen. For every successful query retrieval, the time limit is lesser than the upper bound value.

Figure 1 describes the system architecture consist of five peers, each peer can be connected to all others in the networks. The centralized server database is maintained to store all the common files which are all going to be shared. A single peer having its own memory as well as routing table updated history. The peer can maintain the User Preference Table (UPT) to access the most frequently accessed files more than three times.

A peer can look up the files (records) utilizing Dynamic search technique. It checks the routing table thoroughly, and if any address is accessible, it specifically goes to that peer else it thoroughly searches the network[1].

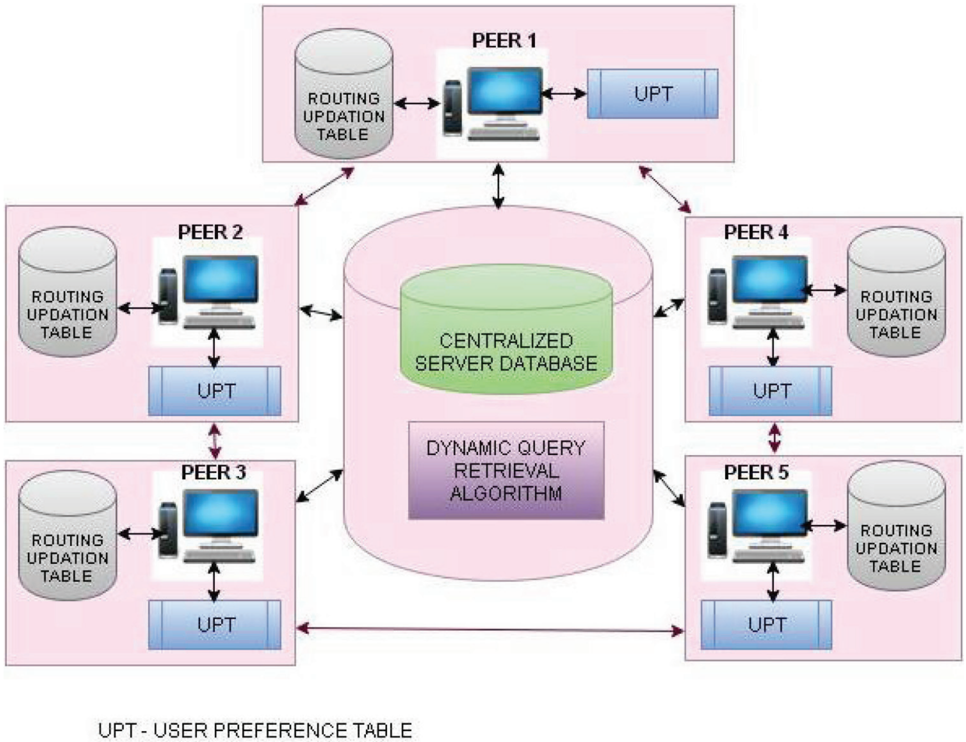

Figure 1. Architecture diagram of proposed system.

It takes the asset from the accessible peer, as well as, at the same time it takes the location of that specific Peer and another name of the sharing resource[1].

2.2 User Preference Model

For many applications, users' various preferences often exhibit statistical patterns of prime importance, which can be efficaciously implemented to enhance the service quality of those applications. This particular segment handles a fundamental issue with a reference to how client's preference can be modeled appropriately. The propounded UPM directs at characterizing users regular interest patterns within the periphery of aprobability-theoretical structure. UPM explains in depth thelog-linear conditional probability distribution, which is denoted as,$Pr(f_d|f_s)$ between any two files f_s and f_d. In this specific propounded approach. The probability representing the user's interest of sharing file f_d, is denoted as $Pr(f_d|f_s)$and the fact that the peer shares another file fs.

$$Pr(f_d|f_s)=\frac{1}{Z(fs)}.2\sum_{h=1}^{k} \lambda h.Fh(fd,fs) \quad (1)$$

Due to its perspective for the modeling conditional probability distributions, probabilistic inference in UPM is very reliable and efficacious in nature, without relying upon any independent suppositions as demanded by other probabilistic modeling methodologies used such as the Hidden Markov Model (HMM). It should also be kept in mind that to propound a standard probabilistic design-which is well suited fora variety of applications,has not been targeted in the earlier researches conducted. For the intended purpose of this specific propounded approach exclusively, we verified that UPM is sufficient to model users' common interests and also todirect the process of discovering the resource and retrieval too. For instance, a same file is searched for more than three times, the details, which were present in the routing table, will be automatically shifted to the User Preference Model Table. Generally, the files which are apportionedvia a P2P network can be unambiguously clarified in-depth with the assistance of a set of attributes.For instance, in a P2P system that shares images, which can be described with the help of features, such as, name, size, memory, and pixel. In this paper, a feature or an attribute is signified byα, and a record or a file, f is further symbolized by a list of features or attributes, f= {α1... αn}.

User Preference Model table contains the information about the past successful retrieval results. It contains name of the peer, address, resource name, and count of the search. UPM table will update every search result. It also has the other sharing resource names. It could help for future reference and it leads the dynamic search. The search based on the user's preference. It will estimate the user's preference between two peers. It gives the priority for the user's preference resources.

2.3 Dynamic Query Retrieval Algorithm

There are numerous search or retrieval algorithms based on knowledge- like Modified Breadth-First-Search, Intelligent Breadth-First-Search, and Distributed Resource Location Protocol, whichare considered to berelevant to combine with the DQR algorithm and any caching or training operations will be benefitting from our DSN algorithm. It exhibits the general plan of incorporating theabove mentioned retrieval algorithms with the Dynamic Query Retrieval algorithm. The probabilistic function depending on the information learnt from the earlier experiences, is constructedin respect of each retrieval time, retrieval target, and local topological information. Hence, a peer has got the necessary information, which is sufficient enough to intelligently adjudicate on the fact as to how many query messages ought to be sent in actual and to which peers these messages, are ought to be directed.

For each neighbor and each item, a probability table is built by the peer, which is applying the DQR system. The probability table is consistently updated by the experiences of the query retrieval. When a search query related to some object or an item gets distributed to a certain neighbor successfully, then probability entry corresponding to that neighbor's object will be increased. At each phase, the retrieval strategy's preference relies wholly on the hop count hc of the query content and the decision threshold 'n' of the DQR. When hc≦n: The number of neighbors which query source sends query messages or content to client based on pre-defined transmission probability p. The DQR pseudo code is represented in below.

Algorithm: The pseudo-code of Dynamic Query Retrieval DQR

```
Input: Query source s, queried resource q, transmission
   Probability p
   Output: The location information of q
   DQR(s, q, p)
   /* the operation of s */
   hc ← 0
   if (hc<= n)
   hc← hc + 1
   s choose p portion of its neighbors
   mi caring hc visits these chosen neighbors
   else if (hc> n)
   hc ← hc + 1
   mi caring h visits one neighbor of s
   /* the operation of r */
   For each (r)
   if (r has the location information of q)
   r returns the information to s
   mi stops
   else if (hc> TTL)
   mi stops
   else if (hc<= n)
   hc ← hc + 1
   r choose p portion of its neighbors
   mi caring hc visits these chosen neighbors
   else if (hc> n)
   hc ← hc + 1
   mi caring h visits one neighbor of r
```

3. Results and Discussion

3.1 Routing Table Updating Protocol

In this specific section, it has been discussed in detail that how a protocol for the updation and the analyzation of the routing tables has been used. Figure 2 explains the flow of routing table updating process. The routing table entry has been updated after every successful query search and retrieval of information.

Figure 2 exhibits the flow as well as the details of our propounded protocol for updating the routing tables and have been described elaborately. Whenever the search operation is carried out by any query $q=<p_q,f_q,h_q,TTL,t_s,t_e>$, and gets to completion efficaciously, subsequently a new entry $E_p=<p_i,f_q>$will be generated, which will exhibit the fact, that the peer 'p_i' will share the queried file 'f_q', which can be temporarily stacked into the routing table entry denoted as R_p. Every peer, p, is stored into the search his-

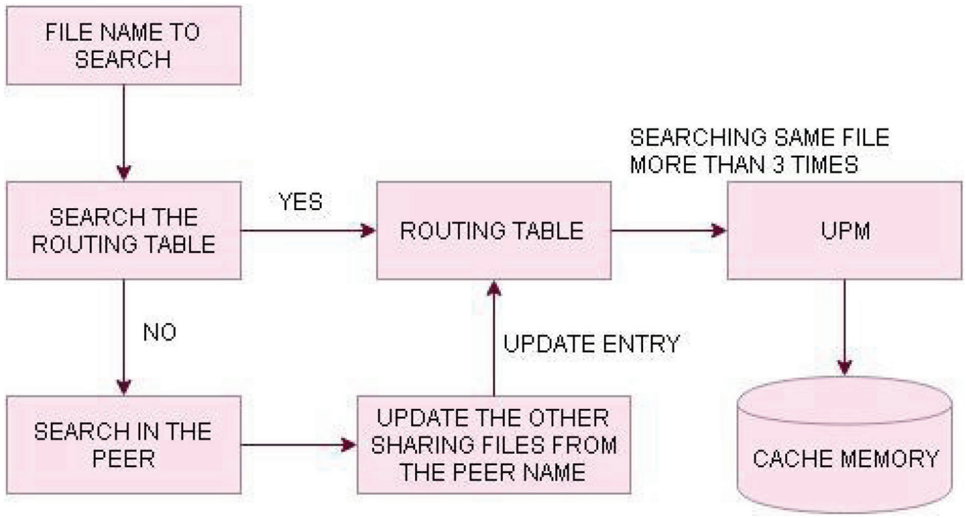

Figure 2. Flow diagram for routing table updating.

tory h_q. R_p is not full, and any other additional entries of R_p will be dispatched.

The major difference between our protocol and other widely used methods for routing table update lies in the way routing entries are chosen for removal. Frequently not used entries are no longer valid in the routing table. For the propounded approach of this paper, the required distance between p1 and p is estimatedin respect of each routing entry E_p=< p1, f1> i.e., maintained by peer 'p'. The probability of removal of any entry is directly proportional to d (p1, p); r. Figure 3 shows search performance under different network settings. In relation to the propounded approach, three challenging schemes or strategies that are to be identified here for updating routing tables are as follows:

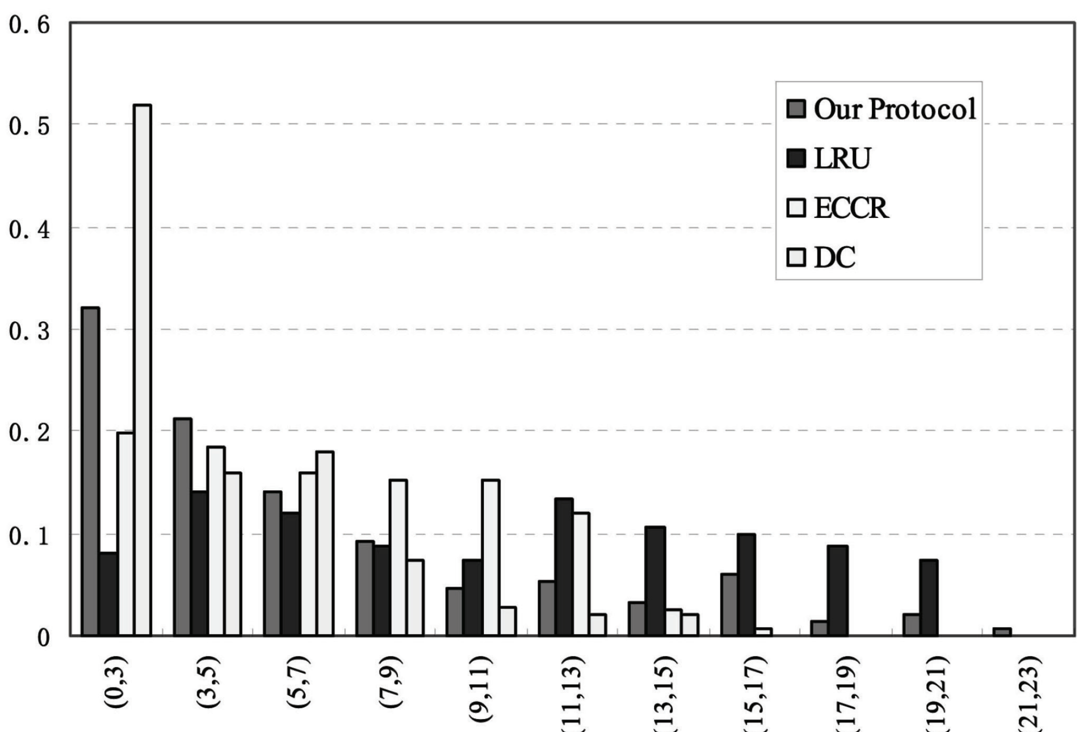

Figure 3. Few observed distributions of neighbor peers with respect to the preference distance.

The LRU (Least Recently Used) scheme:- The routing entry, which is least frequently used to forward the queries, will be discarded from the routing table, without saving the entry in the routing table.

The Distance-Centric (DC) strategy:- The peer p1, which has got the longest user preference distance from another peer p, or else peer p2, the next highest distance from the source peer, will be discarded from the R_p of peer p routing table, depending solely on the probability Pr_d.

The Enhanced Clustering Cache Replacement (ECCR) scheme:- With a particular probability value Pr_e, the least frequently accessed routing entry will be removed from the history automatically. And also the neighbor peer p1, with the longest preference distance from the peer p, will be automatically dropped from the routing entry R_p[9].

4. Conclusion

It has been demonstrated exclusively that the performance of the search process, in the unstructured p2p networks can be efficaciously enhanced by utilizing the statistical patterns over preference pattern of the users', by the propounded approach. The propounded approach has enhanced and developedthe search performance alongwiththe fast retrieval of resource location in thebelow mentionedthree steps: A User Preference Model or UPM has been presented with the aim of capturing users' various preferences within a probability-theoretical structure. This directs to the additional introduction of a concept of preferred distance between any two peers in the unstructured networks. Guided by User Preference Model, a Dynamic Query Retrieval Protocol has been propounded to determine the distributed search of queried files and retrieve resource location through peer's local interactions. Lastly, a protocol for routing table updating can be utilized for the automatic updating and has been propoundedin order tohandle peers neighbor lists and their shared files. To conclude, it is perceived that the solution that has been presented the propounded approach can be extended from many aspects. For an instance, UPM can incorporate additional attributes which will distinguish the network. The live information of

Resources from p2p networks are to be collected and thedata can be employed to examine the efficaciousness of our protocols in the above propounded approach.

5. Reference

1. Bjorkqvist M, Chen LY, Zhang X. Minimizing retrieval cost of multilayer content distribution systems. IEEE International Conference on Communications (ICC); 2011. p. 1–6. PMid:21624468
2. Bukowski A, Nawarecki E, Duda A. Indexing and searching learning objects in a peer-to-peer network. IEEE EDUCON Conference; 2010. p. 1665–70.
3. Chen G, Low CP, Yang Z. Enhancing search performance in unstructured p2p networks based on user's common interest. IEEE Transactions on Parallel and Distributed Systems. 2008; 19(6):821–36. Crossref.
4. Cui Y, Ma T, Cheng X. Multi-hop access pricing in public area WLANS. IEEE INFOCOM Proceedings; 2011. p. 2678–86. Crossref.
5. Dutta N, Kotikalapudi R, Saxena A, Chellappan S. A multi-tiered architecture for content retrieval in mobile peer-to-peer networks. IEEE 12th International Conference on Mobile Data Management; 2011. p. 104–9. Crossref.
6. Jin H, Ning X. Improving search in peer-to-peer literature sharing systems via semantic small world. 15th EUROMICRO International Conference on Parallel Distributed and Network-Based Processing (PDP'07); 2007. p. 31–8. Crossref.
7. Kumar P, Sridhar G, Sridhar V. Bandwidth and latency model for DHT based peer-to-peer networks under variable churn. Systems Communications (ICW'05, ICHSN'05, ICMCS'05, SENET'05); 2005. p. 320–5. Crossref.
8. Anita X, Kumaravel A. Theoretical analysis of trust-based routing schemes for wireless sensor networks. Indian Journal of Science and Technology. 2015 Nov; 8(32):1–4. Crossref.
9. Vijayan K, Raaza A. A novel cluster arrangement energy efficient routing protocol for wireless sensor networks. Indian Journal of Science and Technology. 2016 Jan; 9(2):1–9. Crossref.
10. Li Z, Shen H. Mdr: A p2p-based market-guided distributed routing mechanism for high-throughput hybrid wireless networks. 9th Annual IEEE Communications Society Conference on Sensor Mesh and Ad Hoc Communications and Networks (SECON); 2012. p. 236–44.
11. Shen H, Li Z, Yu L. A p2p-based market-guided distributed routing mechanism for high-throughput hybrid wireless networks. IEEE Transactions on Mobile Computing. 2015; 14(2):245–60. Crossref.
12. Shila DM, Cheng Y, Anjali T. Throughput and delay analysis of hybrid wireless networks with multi-hop uplinks. IEEE Proceedings INFOCOM; 2011. p. 1476–84. Crossref.

13. Shin WY, Jeon SW, Devroye N, Vu MH, Chung SY, Lee YH, Tarokh V. Improved capacity scaling in wireless networks with infrastructure. IEEE Transactions on Information Theory. 2011; 57(8):5088–102. Crossref.
14. Tan Z, Chang G, Cheng W, Gao X, Wang H, Li L. A new routing protocol of structured peer-to-peer overlay networks. 5th IEEE International Symposium on Embedded Computing SEC'08; 2008. p. 297–302. Crossref.
15. Wu Y, Min G, Yang LT. Performance analysis of hybrid wireless networks under burst and correlated traffic. IEEE Transactions on Vehicular Technology. 2013; 62(1):449–54. Crossref.
16. Zhang L, Wang Z, Mei T, Feng DD. A scalable approach for content based image retrieval in peer-to-peer networks. IEEE Transactions on Knowledge and Data Engineering. 2016; 28(4):858–72. Crossref.
17. Zhang R, Hu YC, Druschel P. Optimizing routing in structured peerto-peer overlay networks using routing table redundancy. IEEE Proceedings the 9th Workshop on Future Trends of Distributed Computing Systems (FTDCS); 2003. p. 315–21.
18. Arunachalam AR. Utilizing the amenities of unstructured peer to peer framework in extraction of large scale data clusters. Indian Journal of Science and Technology. 2015 Nov; 8(32):1-6. Crossref.