A Survey on Efficient Opportunistic Overlay Prototype with Reliable Routing & Distributed Resource Discovery Services in P2P Network

Mr. Madhav Gopal Sharma[1] Prof. (Dr.) R. B. Patel[2]
[1]Research Scholar [2]Professor
[1]Department of Computer Science & Information Technology [2]Department of Computer Science & Engineering
[1]Singhania University Rajasthan [2]Chandigarh College of Engineering and Technology, Chandigarh, India

Abstract— The role of P2P network is very major in communication technology is does through multiple cooperating clients. As we know network users are growing in day to day life it does heavy load over the network as well as the traffic patterns are also getting dynamic , it does affect networking performance and quality in services , it require to get split out the network load into multiple server and clients areas , it also does challenges for P2P applications , in this case we need to develop such type of P2P network that can create peer to peer networking in cycle , need to integrate network and application so that servers and clients can handle without networking complexity. Since, P2P network developed at large scale which works in distributed manner. However developed system can have capacity to owing properties of nodes which can serve the networking services by joining and leaving as per on demand manner, since P2P is a dynamic network so that topology is unpredictable, which makes P2P more complex due to find the actual location of a node that may appear one time or may disappear at the next time, under these circumstances, it is very big challenge of peer to peer network to handle distributed networking services. In this paper author try to perform an investigation regarding peer to peer network for different communication aspect apart from this ,author also focus on the technology and working scenario of multiple peer to peer network traditionally invented and used now are days, with their benefits and weakness through multiple repute literatures.

Key words: P2P (peer to peer network), Distributed network, online resource management, Throughput, Efficiency

I. INTRODUCTION

Peer to peer system is very useful and most growing networking technology since it resolving multiple challenging issues as describes in [10, 11] that is time efficient performance, scalability, Easy of maintenance, reliability and usability. To achieve these goals P2P used so many techniques. In this survey

paper one try to focus on this different P2P technology separately to attain these goals, with the help of this paper one can analysis the different aspect of P2P working , its technology and Data sharing policies , P2P provides some special features as “adaptability”, it brings us ease of maintenance and higher the performance by resolving data locality issues, so that such adaptability gives us lots of benefits like related to the data management to achieve data mobility, in the similar way we can have following more such benefits of P2P but apart from this advantages P2P need to improve routing techniques for more quality in communication regarding best uses of data and resources sharing. Following features represent the working traits of P2P network.

This paper investigate the services and performance factors of different traditional P2P based system to identify the weaknesses and improve the underline pitfalls of previously proposed mechanism , in the coming session author will go through the detail study of most common p2p network system even at complexity level , one try to resolve the critical issues to design and propose a new system that can provides much better distributed resource sharing over P2P network along with best routing technique. In next session author also describes the naming, topological, structural and routing feature individually to each and every implemented P2P system. Here table 1 describes the P2P resource sharing feature with quality of services factors.

TABLE 1
P2P RESOURCE SHARING FEATURES

	Performance	*Scalability*	*Maintenance*	*Reliability*	*Usability*
Naming	*Protocols Naming Information and Integration*	*Suggesting Required Namespace*	*Managing Naming Issues*	-	*Enhancing Feasibility*
Structuring	*Local Structure*	*Dynamic Structuring*	*Dynamic Structuring*	*Multi Node Paths*	-
Locating or Routing	*Efficient Routing and location tracking*	*Decentralized Policy*	-	-	-
Data Handling	*Load distribution and balancing*	*Decentralized Storage policy*	*Dynamic relocation*	*Dynamic load balancing*	*Adapt control over data Replication nodes*
Topology Specifications	*Getting optimum links*	*Replication Schemas*	*Re- Structuring*	*Links availability and failure recovery*	-

II. PERFORMANCE SPECIFICATIONS

Following five major features are directly related with P2P system which plays an important role during p2p networking.

- Performance: This is the total time to perform information insertion and deletion operations, this factor also get include the data locality , load distribution among the nodes and data load balancing issues , performance should get measured for efficiency of routing during operational mode.
- Scalability: this is the capacity of the system to remain traceable and locatable with increasing number of users and data nodes, space and time complexity also get measured in this.
- Maintenance: It is always necessary for the system to maintain time to time for better networking services, it includes network topology management, complex operations, data distribution and representation.
- Reliability: It should be measurable as required as possible in includes failure recovery and avoidance scheme, in takes the opportunity to manage failure node detection and their isolation, data replication should be

measurable at every level of operation to provides backup services, it also provides a way to avail data path to increase availability.
- Usability: Easy network accessing at operational level are mostly required feature of P2P networking it come under this category, it also provides different quality services to make networking more efficient and easy, usability provides flexible system with simplicity of user end interface.
- Naming: This term is used for naming space service usually used for tracking and location the address specification.
- Structuring: To define structure of network or type of accessing and upgrading for static and dynamic mode of communication.
- Routing: Routing is one of the major issues in communication; to get efficient routing network specifications has been designed here.
- Data Handling: Data should be access as easy as possible, unique ID assigned and hashing key functions should design and implement.
- Topological Specification: Network topology plays an important role in static and dynamic communication strategies.

III. LITERATURE SURVEY

P2P resource sharing with quality of service specification is a key problem‘s that has been concerned with allocation of computing network resources in such a way that network can smoothly perform its function as an some acceptable performance state [8] . In order to understand the congestion control problem one need to discuss resource factor and its responsibility over communication network that conceptually describes actually how and Where Congestion has been occurred?[14] there are so many situation in the network where the chances of congestion is higher due to the resource capacity and availability that could be strong barrier in order to make communication successful , if one talk about the resource planning for the best network design they first need to manage bandwidth of channel, buffer optimization and management for buffer space computation, functioning of difference processing functions to help node to participate in successful data transmission in highly congested area [11]. Anytime network can cause failure due to unrelavent design of network that affect overall communication therefore network design need to be excellent, it is always possible to get increase the capacity of resource at higher level but it is not sufficient to control congestion error , they need to implement with required specification of network and planning even at low traffic since when such traffic get increased the problem of congestion takes important for the point of view of fairness and higher throughput both are need to be consider as equally important , here one can conclude it as per it definition as ―without having accurate network resource design network always go down as load get heavier[15].

As in [16] Internet services are growing day by day, now it has been working as a great accelerator of today‘s business for void and data communication. In order perform information process there are so many factors have been introducing that consider as a barrio for business information process. In [13] Ahmad S., Gohar N.D., Kamal A. presents a novel approach for dynamic congestion control mechanism for real time streams over RTP, this research presents the mechanism to handle growing network due to the high demand and production of multimedia applications, author suppose the problem due to the massive growth in multimedia oriented application which uses data in the form of streaming like audio, video, etc such application produces big amount of data continuously causes stress over network that result, bottleneck problem on link due to heavy congested network , to address the solution regarding the bottleneck link author proposed a reliable, dynamic congestion control mechanism known as DCCM which works on the principle of RTP for stream delivery data services , author works on the conceptual principle as they convert persistence congestion from transient congestion. To get detect data loss and transient congestion error author design jitter function that performance dynamically therefore network state can also be find at any time with their congestion rate , mechanism take action against congestion error before happening of congestion. Scheme is special works for streams data application, the major object is to identify congestion before it affects and losses data.

Kai Shi,Yantai Shu et. all. in [12] , presents mechanism for wireless network belong to the services category of protocol IEEE 802.11 wide area network. Author design a novel point to point connection oriented congestion control mechanism known as ―Media access congestion control ― the major object behind this research is to control on fluctuate rate of sender and receiver window[17] . MACC improve the performance by having control over sending rate. so that synchronized communication can be achieve to get successful communication in wide area network ,such congestion control has been done at MAC layer to get reduce flow error occur due to congestion, MACC is useful to utilize channel capacity effectively and also capable to manage fairness at both end. Proposed model has been design after the analytical study of wired and wireless communication architecture regarding the study of congestion. In order to eliminate congestion from wireless local area network author mainly focus on the experimental study of congestion at wireless network on TCP reliable congestion less protocol.

IV. RELATED WORK

In this section author discus the traditional P2P network protocol and architecture to get address the performance and problem issues of different system with the help of this we can combine the feature and quality issues of different architecture of this network , at the end we make a single more efficient and routing isolation based algorithm that can resolve problem issues exits previous , to make this possible one need to go through with analytical study of multiple implemented structural and non structural based architecture of P2P network. Let’s discuss the P2P network here.

A. Napster

It belongs to structured based P2P network approach where it works effectively with centralized management system means it provides scalability, reliability, availability and usability form centralized based operation policy. This

protocol works only on specific required features nothing additional feature makes it different but as per the study author find, this system works with many serious issues which is required for better P2P services [18]. It's sort of simplest scheme which work good socially compare to the other since it uses centralized based approach it uses its own namespace for addressing so that in searching operation if a client requesting for a Napster Server, it search all over the client assigned by the one server, if not find address the search to the next server in sequencing manner until it finds the requested client.

The big weakness of this scheme is single node server failure will down the entire network due to centralized server management policy and this model will not replicate the data for backup instead of this uses "keep alive system" [19].

B. Gnutella

It is very earliest P2P system which follow decentralized communication policy , this protocol works with TCP protocol for reliable connection oriented services to make networking and resource sharing more reliable, to get connect with the node of the system one outsider node need to have IP address of at least one node which is already in the system, in this process node will broadcast a "join" message to entire system through participated node , the related node will response against this broadcasted message regarding sharing of storage , files and resources, it uses file name as a key , in order to save bandwidth of system a node will not response against requested query if is not found any thing match with suggested "join "message and if match happen node will return only filtered result set of data [20]. Since it is decentralized system so that it is immune from single point of server failure issues that's why it is better compare to Napster system , but problem which we found in this system is , node are connected loosely which means is it suitable only for small scale network not good for large scale networking.

C. TRIAD

It is a P2P system which is specially designed for content based routing problem issues. The major object of this system is to reduce the content searching time and improve performance of the system , to make it possible it work on the concept of "Network Integrated Content Routing", it put as an intermediate system between centralized and decentralized system since it implement and support replication technology. The specialty of this system is it integrates content router to perform IP routing with name server services. The main concept behind it that content router hold the address and name of the next host so that routing has been performed by server itself, it also offering piggybacking connection so that it can able to easily track and locate the data as required. This system belongs to structured technology so that reliability is increased and performance gets higher as possible due to its name based content routing strategy, therefore it reduces high overheads compare to DNS based system, at the end we can say that the core objective behind TRIAD to integrate content based routing and server to reduce overall network maintenance and overheads so that it can replace the traditional NAPSTER type of system.

D. Pastry

Pastry as described in [4] is a generic P2P content and routing oriented self organizing group of nodes P2P network where node are connecting through internet, this model comes into the category of decentralized ,failure free, scalable and reliable system that routes a message to other participated nodes with some node bits as a key integrated in the message. The major benefit of this system is that it can self detect the arrival, departure and nodes lost issues. Features of this technique are:

- Naming: This techniques uses 128 bit unique node ID, in the process of participation, any node will use this ID to process cryptographic hash with public key, with the help of this naming services node specification nodes are distributed uniformly the set of request as per the node space, the generated resulting 128 bit key will be the original key, to store data in node the ID closest to the Key has been chosen.
- Structuring: The structure of this technique will manage the routing table of node information for the next node and leaf node to perform reliable routing.
- Routing: Here network consisting of multiple of N nodes along with routing table has been organized in logN rows and column in which the last row contain node ID and related information with the help of this we can manage the multiple distributed interfaces to resolve the routing complexity issues and manage the routing table as easy as possible at every distributed area space.
- Neighborhood Node: Space in this one can find the node along with their parent and neighbor nodes , it is designed by the concept of node IDs and IP address are closest to the other required node.
- Leaf Node set: Here set of leaf node maintain the address of node IDs and IP of half nodes who is conceptually closest to larger ID and smaller IDs which is related to the present node .
- Data Managing: This techniques support dynamic data objects to perform run time insertion and deletion on demand basis.
- Topology Updating: It supports dynamic modification in node operation regarding node joining and departure.

E. Plaxton

As in [5] is a distributed structure techniques it is also known as Plaxton Mesh Network since it does support network overlay P2P for tracking naming nodes and routing information to any other object. Plaxton Consist following features:

- Naming: Here node can be identify with knowing their location and semantics , it is using hash technique to mange node random bit sequences , system assumes that the nodes are distributed in node IDs and naming spaces.
- Structuring: In This techniques node uses their memory in two parts, first one manage neighbor table and second one is managing pointer address list, in order to make related multiple run time nodes one need to get point the node to node structure and their services by having multiple route and their management, where the order to proceed this thing is managed by the LogN

complexity. Here point of any node will also manage the copy of object to keep record for future tracking.

- Routing Policy: To process request to it desired location client node will sends message to the destination node it is routed by source node route itself, in any stag if message found relative node, in result it routed to the server directly with object information.
- Topology Updating: These techniques not allowed dynamic insertion and deletion of node; since it is a static type of data structure and therefore it is also get suffer when network gets failure [24].

F. Tapestry

As in [6] it is model after plaxton model, so that it's naming, structuring and routing policies are same as Plaxton but apart from this it has some addition features that it is support it is adaptability, failure recovery which makes it more desirable compare to the Plaxton mechanism.

- Structuring: It is very much similar as Plaxton structure, where each node having their memory structure, divided in two parts to manage point for the next node and pointer address list for the suffix node. Every node also manage a back point list which manage the nodes information where it is pointed as neighbor node, it uses nodes integration algorithm to find neighbor for a specific node. Plaxton assign a dedicated root but Tapestry doesn't do this , it uses distributed algorithm named Surrogate Routing , to continuously create a unique root for an object so that every node having multiple root for the process.
- Routing Scheme: Tapestry perform hashing technique for targeted node ID, generate set of multiple root to process best and search roots over there. Plaxton managing multiple copies of data but here in this it store only the location of closest node.
- Data Managing: Tapestry also support dynamic insertion and deletion of the node at run time as plaxton, but Tapestry also insert and delete the message to multiple roots. Complexity of Plaxton and Tapestry are same.
- Topology Updating: More intelligently Tapestry support management of topology at every stage of nodes , change can made based on the replicated copy of node notification and location based changes are easy in this approach.

G. Chord

It is designed by MIT, it simple and less complex architecture where node can easily insert and delete as per the demand basis, as well as it support fast locating and joining.

- Naming: Every machine having its own m bit ID which has been generated by IP and hash function, every record will manage (k,v) information to design unique key for encryption , this ID will uses to track the location of node.
- Structuring: It supports virtual technique to create a virtual circle of nodes along with their IDs, the circle clock site will be known as successor and remaining known as predecessor node ID. To perform routing efficiently, every node manage part of mapping information, the virtual circle will be part in 1+logN lines as a segment such design occupied only O(logN) memory to manage network information, which support fast routing and locating target machine.
- Locating and Routing: To locate a record with key K in virtual circle naming space should be calculated as P=hash (K).Tracking can start from any node, by information table, where node can find successor and next node.
- Data Managing: To achieve availability, data can be migrated by performing multiple hashing technique, this thing prepare back up functioning at the desired level so network achieving O (logN) time complexity.
- Topology Updating: In Chord, machines can join and leave at any time. For normal node arrival and departure, the cost is O (logN) with high probability, but in the worst case, the cost is O (N). The node failure can also be detected and recovered automatically if each node maintains a "successor-list" of its nearest successors on the Chord ring .

H. CAN

The Content Address Network as describes in [9] is a distributed network infrastructure that support efficient tracking and look up operation, as per the internet scales.

Naming: As defines in [9] CAN, node are identified by their IP address. Each record having their unique ID as a key hash has been functioned as P=hash (K) for every key ID, in this approach the point define the position of node in network for efficient routing.

- Structuring: CAN perform zoning distribution to set of nodes virtually every physical node having their own zone with data record which is mapped by hash function, such node space is known as d-dimensional zone which is partitioned in two parts having next and previous zone detail.
- Locating and Routing: For every generated key, virtual location of node has been computed, the broadcasted query message passed to each and every node of the network until it finds the target machine IP
- Data Managing: CAN support record insertion and deletion at run time, this system has capacity to replicate data for next node, due to its conceptual aspect it's strong and reliable fault tolerance system.
- Topological Updating: As per the rest of the scheme it is also support insertion and deletion of node at run time, fault isolation and maintenance in easy.

V. COMPARATIVE ANALYSIS

In this section author cover the analytical aspect of different P2P systems based on their initial property and feature as defined in table 2 and table 3.

TABLE 2

COMPARATIVE STUDY OF P2P TRADITIONAL SYSTEM

	Napster	Gnutella	TRIAD	Pastry	Plaxton	Tapestry	Chord	CAN
Decentralized	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Space Cost	O(N)	Depends	O(logN)	O(logN)	O(logN)	O(logN)	O(logN)	2d
Data Read	O(1)	O(N)	O(logN)	O(logN)	O(logN)	O(logN)	O(logN)	O(N)
Data Insertion	O(1)	O(1)	O(logN)	O(logN)	O(logN)	O(logN)	O(logN)	O(N)
Data Deletion	O(1)	O(1)	O(logN)	O(logN)	O (log*N)	O (log*N)	O(logN)	O(N)
Node Insertion	O(1)	O(N)	-	O(logN)	Hard	O(logN)	O(logN)	O(N)
Node Delete	O(1)	O(1)	-	O(logN)	Hard	O(logN)	O(logN)	O(1)
Node Failure	-	-	No	O(logN)	No	-	O(logN)	O(1)
Locality	Yes	-	-	Yes	Yes	Yes	No	No
Node Failure	-	No	No	No	No	No	-	Yes

- "-" means data is not accessible.
- "No" means special technique not defined.

TABLE 3

COMPARATIVE STUDY OF P2P TRADITIONAL SYSTEMS BASED ON QUALITY OF SERVICE SPECIFICATION

	Napster	*Gnutella*	*TRIAD*	*Pastry*	*Plaxton*	*Tapestry*	*Chord*	*CAN*
Performance	*Poor*	*Poor*	*Average*	*Average*	*Average*	*Average*	*Average*	*Average*
Scalability	*Poor*	*Poor*	*Average*	*Average*	*Average*	*Average*	*Average*	*Average*
Reliability	*Poor*	*Average*	*Average*	*Average*	*Moderate*	*Average*	*Average*	*Average*
Maintenance	*High*	*High*	*Moderate*	*Average*	*Poor*	*Average*	*Average*	*High*
Usability	*Average*	*Average*	*Normal*	*Normal*	*Normal*	*Average*	*Normal*	*Normal*

VI. CONCLUSION

In this paper author address issues related on P2P distributed resource sharing network, in this paper author analyzed traditional P2P network like Napster, Gnutella, TRIAD, Pastry, Plaxton, Tapestry, Chord and CAN. In which discuss various critical and serious issues based on complexity level as well as performance level, at the end based on the quality of services factors as discussed in table 2 author found that Tapestry, Chord and CAN are some model with good services factors.

In section ii we discussed the introduction part of this survey where author defines the objective of this paper with performance specification as mention in table 1.

In the next section iii we go through the literatures to get related analysis of p2p systems so that at the end we can conclude the survey result.

In section iv we discussed the comparative study of different issues among multiple p2p system with the help of table 2 and 3, here we can find the suitable system as per the complexity level and performance measurement so that at the end of this paper author finally got the poor, average and high performing p2p system to enhance it and investigate it for future research work. But after this deep analysis author found need more suitable resource sharing infrastructure that can perform more better and excellent performance for distributed resource sharing.

REFERENCES

[1] Amna Qureshi et. all, "Stat of the art , challenges and open issues in integrating security and privacy in content P2P system", published in 11th IEEE international conference Digital Information Management,2016 page no. 1-9, DOI 10.1109/ICDIM.2016.7829784.

[2] Zhi Wang et. all,"Strategies of collaboration in multi-swarm peer-to-peer content distribution" , published in IEEE TUP Jounals & Magazines,2012, vol.17,issue 1,pages 29-39,DOI 10.1109/TST.2012.6151905.

[3] Knowbuddy's Gnutella FAQ, http://www.rixsoft.com/Knowbuddy/gnutellafaq.html

[4] Rowstron and P. Druschel, Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. Accepted for Middleware, November 2001. http://research.microsoft.com/~antr/PAST/.

[5] Greg Plaxton, Rajmohan Rajaraman, Andréa W. Richa. Accessing nearby copies of replicated objects in a distributed environment. ACM Press New York, NY, USA Pages: 311320 Series-Proceeding-Article. 1997. ISBN:0-89791-890-8.

[6] John Kubiatowicz, David Bindel, et al. OceanStore: An Architecture for Global-Scale Persistent Storage.In Proceedings of the Ninth international Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2000), November 2000.

[7] Ben Y. Zhao, John D. Kubiatowicz, and Anthony D.Joseph. Tapestry: An Infrastructure for Fault-tolerant Wide-area Location and Routing, U. C. Berkeley Technical Report UCB//CSD-01-1141, April 2000.

[8] Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, and Hari Balakrishnan. Chord: A Scalable Peer topeer Lookup Service for Internet Applications,ACM SIGCOMM 2001, San Diego, CA, August 2001.

[9] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp, Scott Shenker. A Scalable Content-Addressable Network. In Proceedings of the ACM SIGCOMM, 2001.

[10] O'Reilly P2P Directory. http://www.openp2p.com/pub/q/p2p_category.

[11] Peer-to-Peer Working Group. http://www.peer-to-peerwg.org/.

[12] Kai Shi, Yantai Shu, Juan Feng, AMAC layer Congestion Control Mechanism in IEEE 802.11 WLANS IEEE Communication and Networking in Chaina , 2009. pp. 1-5.

[13] E.G. Gran, Reinemo S.A. Lysne O. Skeie T. Zahavi E. Shainer G.,Exploring the Scope of the InfiniBand Congestion Control Mechanism,IEEE Parallel and Distributed Processing Symposium, 2012, pp. 1131-1143.

[14] S.A. Nor, S. Hassan, O. Ghazali, M.H. Omar ―Enhancing DCCP congestion control mechanismfor long delay link", International Symposium on Telecommunication Technologies (ISTT), 2012 pp. 313-318.

[15] M.R.J. Sattari, Noor R.M., Kesshavarz H. ―A Taxonomy for Congestion Control Algorithm in Vehicular Ad Hoc Networks IEEE Journal of Communication Networks and Satellite (ComNetSet),2012 pp.44-49.

[16] O.Habachi, Yusuo Hu., Van der Schaar M., Y. Havel, Wu. Feng,QoE Aware Congestion Control Algorithm for Conversational Services IEEE ICC, 2012 pp. 533-537.

[17] Shi Zhiqiang, D. Ionescu, Dongli Zhang,A TokenBased Method for Congestion and Packets Loss Control Latin America Transaction IEEE, 2013, Vol. 11, pp. 802-811.

[18] F. Paganini, E.Mallada,A Unified Approach to Congestion Control and Node-Based Multipath Routing IEEE/ACM Transactions on Networking, 2009, Vol. 17, pp. 1413-1426.

[19] M.Manjul, R. Mishra, ―A New Equation Based Single Rate Multicast Congestion Control IEEE International Conference on Computing for Sustainable Global Development (INDIACom), 2014, pp. 927-933.

[20] S. Ahamd, N.D. Gohar, A. Kamal.A Dynemic Congestion Control Mechanism for Real Time Streams over RT IEEE Advanced Communication Technology, 2007, Vol. 2, pp. 961-966.

[21] Kai Shi, Yantai Shu, Juan Feng,AMAC layer Congestion Control Mechanism in IEEE 802.11 WLANSIEEE Communication and Networking in Chaina , 2009. pp. 1-5.

