Peer selection technique for efficient searching and streaming in P2P network

Mohanjeet Singh[1], Anuj Kumar Gupta[2]

[1] Research Scholar, IKGPTU, Jalandhar, Punjab, India

[2] Professor, Chandigarh Engineering College, Mohali, Punjab, India

***Abstract*—*significant part of the Internet traffic is generated by peer-to-peer (P2P) applications. P2P could be a promising start for enabling large-scale streaming systems because current internet does not widely support IP multicast and content distribution network is costly. P2P used traditionally for file-sharing butin recent times for real-time communications and live media streaming. Better availabilityof multimedia contents depend upon network topology and Streaming architecture. Peer organization, searching scheme and streaming also play important role in end-to-end distribution. Motive of media streaming is to reduce the latency, QoS and scalability. To achieve these objectives nearest neighbors should be discover fast as possible, searching should be perfect and streaming should be possible with less bandwidth. In this paper several points are diagnosed related to find neighboring peer, streaming and searching the contents in P2P network.*

Keywords*—*Overlay network, Multimedia streaming, peer heterogeneity

I. INTRODUCTION

Distributed data storage and lookup services like Chord [2] and CAN [14] is more efficient when the neighbors on the overlay are near to each other on the primary topology. To offerbetter quality of services, multimedia streaming requires contribution of various intermediate peers [9]. We need new simple system which quickly and accurately find the closest peer by simple probes or by using information obtain from network distance estimation. On way to construct P2P media streaming is Overlay layer and Data scheduler layer. Overlay layer provide partner node and organize the peers for data transferring[11]. Data layer replace the data among peers. Designers of peer-to-peer systems are usually concerned by the match between logical overlay and physical infrastructure. We face a usual challenge of peer-to-peer simulations: as peers are expected to be spread all over the world, the simulator should conform to a model of Internet, although this mapping is known to be a real issue [8]. Existing peer-to-peer applications commonly would take advantage from an Internet similar the original network, wherever theselimitations were not in place [21]. Some previous research assume that it would be convenient to estimate the closest peer for newcomer accurately and quickly if exact route between each peer and central server can be store by the central server itself. The reasoning behind such assumption indeed refers to the statistical regularities observed in the large -scale structure of Internet [9].

Most important multimedia applications like video on demand, IP-TV, video streaming are based on P2P network. P2P network provides the structural possibility of mutually sharing and distributing user-created contents video and music in a variety of user community environments. These networks aggregate large amount of heterogeneous nodes known as peers. Thesepeershavesomeinteresting characteristics like self-configuration, adaptation and self-organization. Multimedia streaming over internet is managed by content distribution networks (CDN) [7] such as Akamai [1], Limelight Networks [17]. CDN contain dedicate servers which perform the functions of content storing and serving the client demands. Multimedia streaming needvery high bandwidth.The server generates a multimedia stream which is divided into blocks and delivered to a small subset among the participating peers. All peers exchange these blocks to produce the stream again. Those peers which are concerned may have heterogeneous upload bandwidth capabilities. Some problems has been already discussed in ([9], [10], [4], [18]) which shows that P2P multimedia steaming is still in its growing stage and for improvement in P2P multimedia streaming important research and investigations efforts are required. Objective of this paper is to provide basic knowledge of peer organization, multimedia streaming over P2P and consider the issues related to this area. Paper is divided into six sections. P2P overlay classification is describe in section II, which discuss the type of Peer-to-Peer network, Different scheme to select nearby peer in section III, Which discuss various schemes for selection of nearby peer, P2P multimedia content access in section IV, Searching schemes in section V, which describe different searching schemes used to search semantic contents in P2P network, multimedia streaming architecture in section VI, Multimedia streaming on p2p networking section VII.

II. P2POVERLAY CLASSIFICATION

A. Unstructured

In this type of network as the name indicates they do not maintain any specific structure on the network topology As such, there is not much systematic information to help index objects across nodes. So nodes often locally maintain objects they share, and search in the network is more or less a blind process, passing queries around nodes to check whether they have the desired objects. Since search is done locally, complex queries like wildcards can be supported as they were in a centralized system. Search space grows in proportion to recall rate because search is a basically a blind process. [12]

B. Structured

These networks areso named becausethey maintain a structure or topology to maintain a network. Most Structured peer to peer operate using distributed hash tables (DHT) where given a key, they can efficiently locate an object having the key by just maintaining neighboring tables entry. Nodes cooperatively maintain routing information about how to reach all nodes in the overlay [4]. Structured overlaysprovide a limit on the number of messages needed to find any object in the overlay which is particularly important when searching for infrequently occurring or low popularity objects. Local routing table which is there with each peer is used by the forwarding algorithm. When the peer joins the overlay using a specified bootstrap procedure, the peer's routing table is initialized. Peers periodically exchange routing table changesas part of overlay maintenance. Peers who are neighbors in the overlay can be distant in the underlying network because the address space is virtualized and peer addresses are typically randomly assigned.

C. Hierarchical

A hierarchical overlay is an overlay architecture that uses multiple overlays arranged in a nested fashion, and the nested overlays are interconnected in a tree. Hierarchical overlay lookup have two steps higher level routing and lower level routing. The higher level routing isuse to locate the destination domain and lower level routing is used to locate the destination node. To perform the task of higher level routing or lower level routing a node has two ID: a group ID and a node ID. Routing algorithm of underlying DHT determine the lower level routing. Super-node who knows the inter domain routing information can implement the higher level routing. [20]Proposed open sourcesimulator which work in hierarchical environment. It simulates P2P network and open source. It is free to other developers for add or modify networks. SHPSIM is event driven and able to execute number of events.

D. Federated overlay

A federated overlay is an overlay that is formed from a collection of independent overlays, each implemented by a separate administrative domain, and which may use different routing algorithms and addressing mechanisms in each domain. Each overlay is autonomous, andpeering arrangement is required between overlays for messaging operations. Each domain manages some management tasks like authentication, authorization and other for its overlay.

E. Semantic Overlays

In semantic Peers are clustered according to content. These clusters overlap, because peerscan contain different content and belong to several clusters. Query, coming in the network, is spread to related clusters only and flooded among relevant peers. So, clusters, irrelevant to query, don't receive any messages. Each Semantic Overlay Network representsvirtual, abstract and independent layer of previously clustered, classified peers. Such networks play roles of mediators between queries and certain peers, they are responsible for "understanding" the meaning of query, establishing semantic relations between query and peers and implementing query routing to relevant peers and, finally, they significantly reduce over flooding of physical network.

III. SCHEME TO SELECT NEARBY PEER

A. Hierarchy based scheme

The hierarchy is based on topological clustering of these peers, where nearby peers are grouped into the same cluster. The querying member (termed query-host) finds its closest peer by successively refining its search in a top down manner over this hierarchy. Tiers create a hierarchy of the application peers. In this scheme, each application peer dynamically discovers a few other application peers, and is required to make distance measurements to a subset of them. Arrange the set of application peers into a hierarchy. Logically, each peer keeps detailed state about other peers that are near in the hierarchy, as well asonly hasrestrictedawareness about other peers in the group.[15,16] The hierarchical structure is also important for localizing the effect of peer failures. Peers that are "close "with respect to the distance metric are mapped to the same part of the hierarchy. The closest peer-finding operationproceeds stop-down on the hierarchy thus successively refining the search at every step, until the suitable peer is recognized. Hierarchical routing in the Internet offers several benefits over non-hierarchical routing, as well as scalability and organizationalautonomy (e. g., at the level of a university, a corporate campus, or even the coverage area of a base station in a mobile network).

B. *Coordinate-based schemes*

Its aim is to fix the location of any host on Internet, to allow two peers to estimate their latency by a basic distance computation [9]. Previous works show that these virtual coordinates can be obtained by active probing, i.e. by collecting round-trip-time (RTT) measurements between peers and a small set of landmarks [9]. Unfortunately, network coordinate systems require a substantial amount of time before to deliver accurate information and do not take physical network topology into consideration. It is assume that if a central server can store the exact route between each peer as well as itself, it would be capable toguessexactly and quickly the closest peers of any newcomer. The requirement of coordinate based schemes is to install a handful of globally distributed "landmark" servers. Peers are made to open TCP connections with each landmark by adding the addresses of landmarks to the first list of peers returned by the tracker. The measurements obtained from the TCP connection are used to compute the location of the peer in a centralized network coordinate system (GNP)[9].The tracker can thus reply to any subsequent request from that peer with a list of carefully selected peers in place of the usual random choice. Each time a peer P connects to the tracker, the tracker deliver back a set of other peers which exhibit low latency routesto P. The client remains totallyunconscious of this bias and behaves as usual.

C. Physical Topology based scheme

Most traditional P2P media streaming systems don not consider the physical network among thepeers, while selecting partner candidatesaccording to physical network can reduce the latency of the data transferring and improve the quality of service. How to decide the peers which are actually close to each other in physical network? Several schemes havebeen proposed to estimate the internet distance. Internet Distance Maps (IDMaps) [9] places tracers at key locations in the Internet. These tracers calculate the latency betweenthem and advertise the measured information to the clients. Clients acquire the latency using the nearby tracer, and can compute the distance between them without extra probing in particular way. GNP (Global Network Positioning) [9] and its sequel NPS (Network Positioning Systems) [7] are another methods for estimating the network distance. In physical topology based scheme there will be a Media Server which is the data server which pushes the data to clients as bootstrap. Index Server stores the landmark list. The client with a tree icon is landmark node who manages sends probing packets (Ping-Pong Msg) and records the RRTs value from the landmark nodes. Once p receives the response from the landmark nodes, the latency from landmark nodes of p make up the vector. The vector is stored on landmark node. Landmark node manages the entire vector in the cluster and returns partner candidates to p according to specific algorithm.

D. Landmark nodes scheme

Landmark nodes measure RTTs among themselves and use this information to compute the coordinates in a Cartesian space for each landmark node. New peer join the system to measure the distance with landmark nodes and compute the coordinate for itself according to thecoordinates of landmark nodes. The Euclidean distance among nodes in the Cartesian space is directly used as an estimation of the network distance. Since any client who joins the system should contact IndexServer initially, Index Server measures the peer information like online time, bandwidth and so on. The server selects the peer with long online time and high bandwidth as landmark nodes. In case the landmark nodes quit, there is a backup mechanism. Before landmark node quits, the node should inform Index Server to update landmark list. In order to increase the system stability, it is necessary to deploy some landmark nodes in advance. As explained in the former section, landmark node is also the leader of the cluster. Every landmark maintains a one hop route table between them in order to search the neighboring peers from one cluster to another. The route table is also stored on Index Server. If single route table modify, it should be updated to other landmark nodes.

IV. P2P MULTIMEDIA CONTENT ACCESSES

P2P content access and delivery has become one of the most popular P2P applications because of high scalability and low cost implementation. This includes P2P sharing of music,video, P2PTV, P2Pradio and P2Pvideo streaming, etc. Content can be delivered viadownloading or streaming. Media stream is segmentedinto data blocks and delivered via flooding or specific route defined by the topology in P2P overlay network. We will discuss multimedia content access in two parts searching schemes and streaming.

V. SEARCHING SCHEMES

Content should be search before it is accessed. Peers are distributed in P2P network there forcontents are scattered and duplicated in a distributed fashion. Content retrieval in a P2P network needs to contemplate the specific network model as well as the characteristics of the content being accessed. Search algorithm should comprise support of complex queries, low cost in implementation, and fast and high accuracy query return capabilities. In structure P2P network static key and ID based object lookups are supported but in unstructured complex queries can be handled.

A. Content Indexing

Content searching is dependent on content indexing. In some cases index is kept in centralized location which is called centralized indexing, searching is done by forwardingquery message to the centralized indexing server to find the location. Location is send back by the server and data is transmitted in P2P fashion.

B. DHT scheme

Structured P2P network widely used DHT based search. DHT requires considerable effort due to the dynamics of the network topology. This scheme relies on numerical keys to index and query objects. Searching is done using key distance and routing towards the peer which has the closest key to the querying object key. DHT is unable to support complex queries.

In the case of centralized indexing searching is efficient and fast but in non-centralizedindexing scheme searching contain high cost with query flooding. Distributed index causes large amount of data being transmitted over the network. To solve this problem conventional informational search scheme can be used. In [5] content summary based inverted indexing was proposed in which bandwidth requirement can be reduced for query flooding becausequery can be transmitted in smaller candidate list.

VI. MULTIMEDIA STREAMING ARCHITECTURE

P2P streaming architecture states the methods used for multimedia content transfer. It also state the entities involved during the streaming. Network structure remains roughly stable throughout a transmission, in terms of diameter, average shortestpath and maximum degree. In contrast, the network clustering coefficient tends to decrease with time, possibly in response to changes in the partnerships of individual nodes.[6]In the case of P2P streaming peer can be a source, destination or intermediate. In the case of source it contains the media contents and share with other peers [13]. Peer can store the full content or part of a given content. If peer play the role of destination peer then it is the client who make the request for content. It can get contents from one or more senders. In the case of intermediate, peer will receive content and transmit it to the next peer. It will work as a transport node. Multimedia contents are distributed using tree structure of overlay network. In the case when contents are store in multiple sources then contents can be either replicated or can be split and dynamically placed in many peers. If contents are replicated then any content can be found in several emplacements into the network. In second case analyses of client request is required to place the different pieces of content in the network. In single source, multimedia contents are stored into single source peer. That source starts transmitting the contents to all clients' peers who are requesting for it. Intermediate peers store some part of the content in their buffer which can be directly retrieved by the new client peer when it joins the network.P2P can be unstructured which do not keepany specific structure on the network topology As such; there is not much systematic information to help index objects across nodes. So nodes often locally maintain objects they share, and search in the network is more or less a blind process, passing queries around nodes to check whether they have the desired 0bjects. Search space grows in proportion to recall rate because search is a basically a blind process. Structured sustain a topology to maintain a network. Most Structured peer to peer operate using distributed hash tables (DHT) where given a key, they can efficiently locate an object having the key by just maintaining neighboring tables entry. Nodes cooperatively maintain routing information about how to reach all nodes in the overlay [3].

VII. MULTIMEDIA STREAMING ON P2P NETWORK

Streaming refer to the delivery method. Numbers of digital video being streamed over the Internet each day are all growing exponentially. This obviously is placing an intense demand on the network bandwidth at the Internet backbone as well as on the servers that are offering the digital video and audio services. Multimedia streaming over internet is managed by content distribution networks (CDN) [7]. CDN contain dedicate servers which perform the functions of content storing and serving the client demands. Dedicated servers provide not only better distribution of files but also better streaming of real-time media. Since servers are geographically distributed therefor it reduces network congestion and better serve clients in given regions with low latency. Servers placed at the edge of the network and closer to users, better quality of experiences can be expected for real time media streaming. Streaming applications can be one-to-one, one-to-many as well as many-to-many. Live personal video is an example of one-to-one or one-to-many. Internet video is an example of one-to-many and video conferencing is an example of many-to-many. According to the streaming application protocols can be unicast, broadcast or multicast. In a P2P network, data can be streamed through a tree specific, mesh specific or multicast overlay.

A. *Tree based streaming*

In tree based approach content rooted at the source node, is pushed along the tree to the destination peers. If any member leaves the tree then tree is broken and children of departure node rejoined to the tree. Tree specific system is unstable and does not utilize the bandwidth of leaf node. Multiple trees may be built to improve the fairness in resources.

B. *Mesh based streaming*

Mesh-based overlays implement mesh distribution graphs for content streaming. Each new node first obtains a content block availability map. It lists the peers who have the desired contents blocks. New node then contacts a subset and request for streaming and obtains the content block.

C. *Multicast streaming*

When there are multiple clients (receivers) simultaneously requesting/receiving the same media stream in a streaming application, multicast can be implemented. Multicast is a special type of streaming where protocols are defined to deliver a packet to a group of destinations at the same time using effectualpolicies. Multicast can be installed at different network layers. IP multicast which implements multicast at the IP routing level is generally high in implementation cost. P2P overlay multicast was invented to reduce deployment cost and improve scalability. AP2P overlay multicast system should implement [3]

VIII. Conclusions

Within coming years it is expected that multimedia services will grow up appreciably. For betterstructure of peer-to-peer system, peers should preferentially be connected with the peers that are the closest in Internet. However determining these closest neighbors in a wide population spread on a large-scale network as Internet is still a problem. According tothe technology, there isvarious technical problems still need to be resolved before P2P network takes on a full spin in multimedia content delivery. An effectual system that can getbenefit of the P2P network resources in a fair and balanced way can have a strong impact on system scalability and performance. The main motivation behind our work is to discuss issues regarding peer selection, multimedia searching and streaming in P2P network so that better designing can be made for reliable solution.

Acknowledgment

I would like to thanks Punjab Technical University Jalandhar, for providing me opportunity and technical support to complete this research work.

References

[1] Akamaihttp://www.akamai.comJ. Breckling, Ed.

[2] I. Stoica, R. Morris, D. Karger, M. FransKaashoek, and H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In Proceedings of SIGCOMM, August 2001.

[3] J. Buford, H. Yu, and E. K. Lua (2008), "P2P Networking and Applications", Morgan Kaufmann.

[4] J. Buford, H. Yu, and E. K. Lua (2008), "P2P Networking and Applications" Morgan Kaufmann.

[5] K. Yang and J. Ho, Proof (2006), "A DHT-Based Peer-to-Peer Search Engine", Proceedings theIEEE International Conference on Web Intelligence, Hong Kong: IEEE Computer Society. 702–708.

[6] Kenia Carolina Gonc¸alves, Alex Borges Vieira, Jussara M. Almeida, Ana Paula C. da Silva, Humberto Marques-Neto, S´ergio V. A. Campos (2012), "Characterizing Dynamic Properties of the SopCast Overlay Network", 20th Euromicro International Conference on Parallel, Distributed and Network-based Processing

[7] Limelight networks http://www.limelightnetworks.com.

[8] L. Dall'Asta, J. I. Alvarez-Hamelin, A. Barrat, A. Vazquez, and A. Vespignani. Exploring networks with traceroute-like probes: Theory and simulations. Theor. Comput. Sci., 355(1):6{24, 2006.

[9] M.Singh, D.S Dhaliwal. "Searching and streaming of multimedia content in p2p overlay network", International Journal of Computer Engineering & Technology Volume 3, Issue 2, August (2012), pp. 433-438.

[10] M.Singh, D.S Dhaliwal. "Multimedia Streaming in Peer-to-Peer: Approaches and Challenges", International Journal of Emerging Technologies in Computational and Applied Sciences Issue 3, Vol. 1 & 2 Feb-2013.

[11] M.Singh, D.S Dhaliwal. *"Peer organization in structured P2P network"* In International Journal of Software and Web Sciences Issue 6, Vol. 1, 2013.

[12] M. Mushtaq, T. Ahmed, D.E Meddour, "Adaptive packet video Streaming Over P2P NETWORKS", Proceeding InfoScale '06 Proceedings of the 1st international conference on Scalable information systems Article No. 59 ACM New York, NY, USA.

[13] M. Zhou and J. Liu, "A Hybrid Overlay Network for Video-on-Demand", IEEE International Conference on Communications (ICC'05), Seoul, Korea, May 2005

[14] S. Ratnaswamy, P. Francis, M. Handley, K. Karp, and S. 'Shenker. A scalable content-addressable network. In Proceedings of SIGCOMM, August 2001.

[15] T. S. E. Ng and H. Zhang, "Predicting Internet network distance with coordinates-based approaches," in Proceedings of the IEEE INFOCOM 2002, 2002.

[16] T. S. E. Ng and H. Zhang, "A network positioning system for the internet," in Proceedings of USENIX Annual Technical Conference, 2004.

[17] V. N. Padmanabhan, H. J. Wang, P. A. Chou, "Resilient Peer-to-Peer Streaming", IEEE ICNP 2003, Atlanta, GA, USA November 2003.

[18] Xiaofei Liao, Hai Jin, Yunhao Liu, Lionel M Ni, and Dafu Deng, "AnySee: Peer-to-Peer Live Streaming", To appear at IEEE INFOCOM 2006, Barcelona, Spain, April 2006.

[19] Yuh-JzerJoung (2007), "Wildcard Search in Structured Peer-to-Peer Networks", IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 11, pp. 1524-1540.

[20] Guruprasad Khataniar, "SHPSIM: A Simulator for StructuredHierarchical Peer-to-Peer Protocol" International Journal of P2P Network Trends and Technology (IJPTT) volume 4 January 2014.

[21] Tiruveedula GopiKrishna, "Future Research Challenges of Peer-to-Peer Systems", International Journal of P2P Network Trends and Technology (IJPTT) volume 8 May 2014.