

P2P Based Fast Replica Algorithm for Semantic Overlay Networks

Kirubakaran. S, Arunkumar. M, Prakash S.P

Abstract: Nowadays the Overlay networks are emerged as a prevailing and bendable technique. This paper aims to optimize the throughput parameter of overlay networks. A network based on similar content is created and called as Semantic overlay Network. To duplicate the contented within the semantic system, Fast Replica methods are used to condense the contented relocate time. This file are down loaded through various network in parallel manner. The vibrant main schemes are obtained to the main end users of the domain SON at the similar time, alluring dissimilar portion of that file from dissimilar nodes and saturated in local point. The loads are animatedly collective in the middle of all the peers. while considering a peer-to-peer (P2P) architecture queries are propagated to Overlay Broker (OB). The OB in turn returns the IP address of the semantic group requested by the node. Then the node queries the semantic members for the file. All the connections are maintained by nodes. For instance, "Jazz" archive of the main node will connect to additional similar nodes. All the nodes of Semantic Overlay Network (SON) are organized by themselves to the OB. The uncertainty are in retreat to the suitable SONs, which increasing the probability of matching files and dipping the look for the dissimilar nodes on the main content.

Keywords: Replica algorithm, overlay networks, IP address, P2P

I. INTRODUCTION

Content Delivery Network (CDN) has mainly to optimize the deliver specific content, such as transaction-based Web sites, static Web pages, streaming media, software packages or even real time video or audio.. There are two universal approach to build CDN [1]. One is overlay model, e.g. Akamai.

The other is the Network model, e.g. IP Multicast. A few content delivery designs are use both overlay approaches and the network. Generally CDN uses the overlay protocol. The overlay system is formed to deliver content to a distributed nodes near the main node. The system communications are formed in the top sheet of the organization [1]. All the network contains network contains different components, such as QoS, IP routing. This paper focus on the creation of overlay network protocol based on semantics, an algorithm FastReplica for replication of content and an approach of Digital Fountain to avoid retransmission while downloading.

Manuscript published on 30 December 2019.

*Correspondence Author(s)

Kirubakaran. S, Assistant Professor Department of ECE, Bannari Amman Institute of Technology-Sathy

Arunkumar. M, Assistant Professor Department of ECE, Bannari Amman Institute of Technology-Sathy,

Prakash S.P, Assistant Professor Department of ECE, Bannari Amman Institute of Technology-Sathy

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

A. Challenges and Opportunities

The main abstract problems in the environment of Internet is node position. [1] The content are delivered to infrastructure must cope with which includes:

- **Asynchrony:** specific receiving end is close and open to maintain or rejoin and leave the infrastructure at any time.
- **Heterogeneity:** The speed and loss rates may vary with respect to time scale.
- **Scalability:** The service are to be scale to the receiver populations
- **Transience:** The most of the routers, end schemes and links may fail, or their recital may fluctuate.

These problems can be overcome with the creation of overlay network.

II. RELATED WORKS

John W. Byers et al., [2004] proposed a new model to dependable distribution of bulk data in adaptive overlay networks. This study represented the optimize of large opulently connected multi layer point in the overlay networks, spotlight on the question of what to put in each transmitted packet. Furthermore the idea of digital fountain with removal codes was established which allows the node to reconstruct the main original node of size n from a subset of any n nodes to various universe of decoded cryptogram. Such approaches afford dependability and a considerable degree of application-level liveness, as it flawlessly tolerates packet loss, parallel node transfers and link node migration.

Daniel Ellard et al., [2004] describes DISP, the Distributed in sequence storage space Protocol. DISP is a sensible, well-organized, protected and fault-tolerant client/server protocol for dispersed data storage. DISP is well-organized in terms of network traffic; even in the corporation of failures, the protocol needs transferring only slightly additional than L bytes in order to recover an object of size L. The DISP is well secure because all of the data is stored and reassign in an encrypted form only. The concession of the network or a server divulge none of the data. The DISP is also able to make certain truthfulness of the data so that the data can be repossess correctly even when servers have been besmirched. Alessandro Falaschi et al., [2004] exemplify the live watercourse of media in contented delivery systems where multicast of information is practiced. By deploying an array of relays along the network, an application level multicast content delivery network can be created. It shows the way, which relay is suitable to serve the client.

It routes the content to the nearest relay of the user. Yan Sun et al., [2003] gives a survey of P2P networks. The node dispute that ought to an overlay to be built on a physical organization in order to progress the presentation of floods and random walks. The keep the content assignment and innovation instrument to complex queries. A new system is introduced to named Structella. It diminishes preservation transparency and improves content exposure presentation, but it does not use structure to systematize the comfortable in the overlay for behind composite queries. John Risson et al., [2004] provides various search methods in a P2P network. It also specifies the methods to retrieve a document based on a keyword search. It tells a mechanism of how queries can be classified and how indexing is done to documents. Richard T.B. Ma et al., [2003] discover a preparation plan for file transfer overhaul. The P2P nodes offers the main documents of larger size. The allocate convey ability of all the minimum and maximum node. Typically the request are purpose of the demand node's connection type, its efficacy purpose the new node of all documents. It converse on how to make capable use of P2P and to supply both equality and inducement to main documents in the P2P community of the social medium.

Yatin Chawathe et al., [2003] proposed the the decentralized and centralized network. The napster lead the idea of peer-to-peer file sharing, and hold up it with a federal file search ability. The P2P structure like Gnutella adopts decentralized search algorithms. He suggests alteration to Gnutella's design that energetically become accustomed the overlay topology and the pursue algorithms in order to put up the normal heterogeneity present in most peer-to-peer organization. The Napster was introduced the first system to recognize sent to a middle server but as an alternative, that already hold the satisfied. The important file-sharing systems are self-scaling in that as additional peers join the organization to look for archive, they have add to the collective download ability. But federal systems have been restoring by novel decentralized systems such as Gnutella, which deal out both the download and search capabilities. The doubt are not sent to a middle site, but are as an alternative distributed in the midst of the peers. Gnutella, the initial of such systems, uses an formless overlay network in that main document is connected and assignment of files within it is mostly unimpeded. It deluges each query crossways this superimpose with a incomplete scope. Upon in receipt of a query, every peer sends a catalog of content identical the query to the invent node. The load of every one node grows linearly with the whole numeral of query, which in twist cultivate with organization amount, this come up to is obviously not scalable.

Pablo Rodriguez et al., [2002] represent parallel admission to multiple mirror location. They proposes two diverse method: 1) dynamic TCP and 2) TCP is side line problem. A parallel admittance, a is clearly transmitted the date to the nearest node from the server. e.g., server single sends the initial semi of the file, and two main nodes are sends the second half. A lively parallel right of entry, a client divider a text into a large numeral of small blocks. The patron first requirements a dissimilar block from each server. The server terminate broadcast a block, the client matter a novel demand for a wedge that has not yet been

requested from any other server. The identical process is frequent for each block awaiting all blocks are completely established. A lively parallel right of entry can with no trouble adapt to varying system/server circumstances. Mohammad Malli et al., [2003] represented the duplication of digital satisfied in overlay networks. A client along with TCP reduces the content transport time significantly. from the most excellent server, or in parallel on or after a division of servers at the summit of the catalog (most excellent servers). Chunqiang Tang et al., [2003] presented a search apparatus for Peer-to-Peer (P2P) organization. The conventional approach has either been federal or uses flooding. He explained pSearch, a decentralized non-flooding P2P in sequence repossession system. The pSearch distributes text directory from side to side the P2P network based on text semantics make by Latent Semantic Indexing (LSI). The look for cost (in terms of dissimilar nodes investigate and data transmitted) for a known inquiry is thus reduced, since the directory of semantically related documents are likely to be co-located in the system. The pSearch is achieving presentation similar to central in sequence repossession systems by penetrating only a small number of nodes. Ludmila Cherkasova et al., [2003] introduced an well-organized and consistent imitation of big files in the. The algorithm represent that to replicate a large sleeve in the middle of n nodes, the unique file is partition into n subfiles of equivalent size and each subfile is relocate to a different node in the collection. Subsequent to that, every node propagates its subfile to the outstanding nodes in the group. Thus as an alternative of the archetypal replication of an whole file to n nodes by means of n Internet paths, between the original node to the replication group. It is straightforward and low-cost. It does not necessitate any change or adjustment to the obtainable Internet communications, and at the same time, it appreciably decreases the file duplication time.

III. SEMANTIC OVERLAY NETWORK (SON)

The node connections in the network are influenced by the content, the nodes are having many new logical connection connected in parallel manner. The server Semantic Overlay Network (SON) are connected to all the node parallel manner [2], [3], as shown in Figure. 1. The nodes C,A,B are clustered jointly. The important nodes such as C, E, F are clustered jointly. The main and nearest Node C is a affiliate of both the groups. The routed uncertainty are connected to the suitable SONs. It is the rising the probability that corresponding files will be establish rapidly, and tumbling the load on nodes are connected parallel manner that have unrelated contented. With Semantic Overlay Networks (SONs), are connected with semantically alike contented are "clustered" jointly and a node may be a associate of additional than one come together. All the important node choose which SONs is to join based on the organization of its documents. The important nodes are connected to issues a query, primary it categorize it and drive it to the suitable SONs [5].

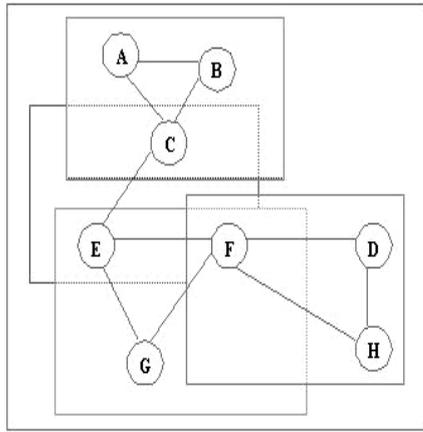


Figure 1. Semantic Overlay Network

IV. FASTREPLICA ALGORITHM

A narrative method named as FastReplica is planned for an well-organized and dependable imitation of great files in the SON [6]. There are a small number of essential ideas subjugated in FastReplica. All the important files are transferred from the various node to all other node present near the server. The each node promulgate the important nodes of the various node in same condition. The characteristic duplication are substituted of an whole dossier to n nodes. The various algorithm are proposed in different modules using n Internet pathway, between the original node to the duplication group, FastReplica are develops $n \times n$ Internet conduit inside the duplication group are used to transferring $1/n$ -th of the main file. Thus FastReplica algorithm container is used for replication of all maximum files to a maximum collection of nodes.

A. UnDistribution Step

From the Figure. 2. All instigator nodal point N_0 undo n concurrent system relatives to nodes N_1 to N_n , and propel to all beneficiary node N_i ($1 \leq i \leq n$) are represent as follows.

- All the important nodes are of the form $\{N_1, \dots, N_n\}$ to which subfile F_i is to be drive on the after that step.
- Subfile F_i .

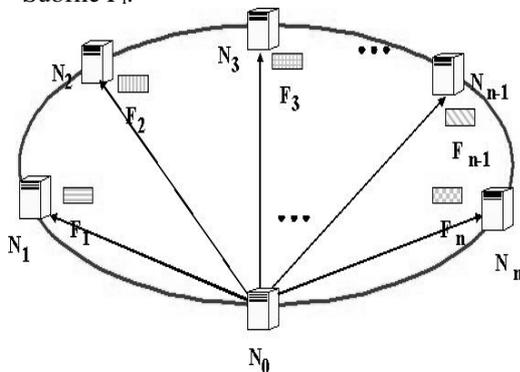


Figure 2. FastReplica -Un Distribution step

B. Collection Step

following receiving file F_i . Figure. 3. shows how node N_1 gather diverse part of file on or after additional nodes.

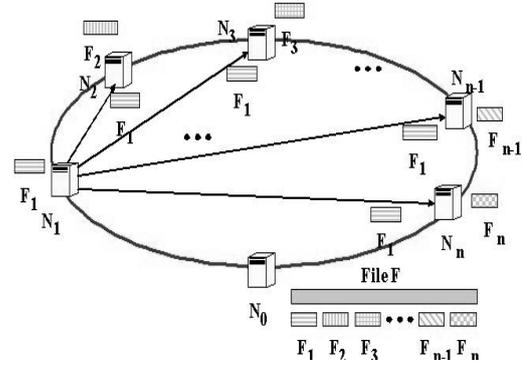


Figure 3. FastReplica - Collection step

C. Resilience Step

If node N_i be unsuccessful when acting as a recipient node all through the distribution step then node N_0 is conscious of the node N_i failure. The corresponding main Node N_0 are used in the subsequent node pair executes the subsequent repair step: The $k - 1$ are opened previous connections to the rest of all different the nodes to send the missing F_i file to both nodes in the node is shown in Figure. 4.

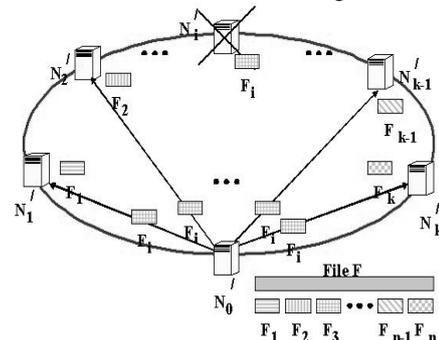


Figure 4. FastReplica - Resilience step

V. OVERLAY BROKER (OB)

The Overlay dealer complex are an significant fraction of an occasion based middleware. In this fraction, we examine the supplies of cover broker networks and disagree that using peer-to-peer method for their formation and the satisfied based routing.

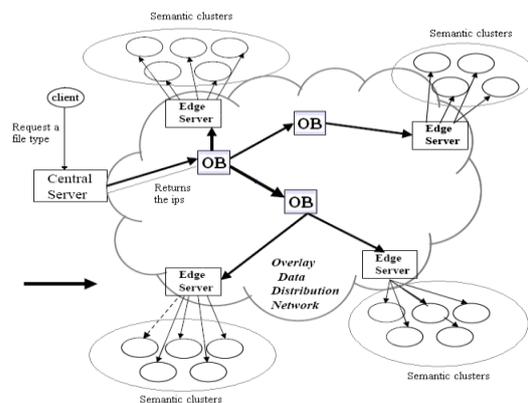


Figure 5. Structure of Overlay Broker

Publish/subscribe communication is becoming ever more significant in the expansion of large-scale dispersed systems. approximately all conventional synchronous, demand/reply middleware display place that are used these days also give a little form of the asynchronous messaging that lend a hand to build systems out of loosely attached components subsequent the circulate/subscribe example. additional middleware display place, such as JMS and web armed forces are muscularly based on asynchronous messaging . It will be the edifice blocks for the subsequently generation of internet-wide, ever-present applications. In such a globe, scalability is dominant and must be ensuring at all coating of the arrangement. The obtained message middleware may potentially have millions of active clients and, consequently, must itself be execute in a dispersed fashion. The effective Content-based routing has been recognized as a commanding idea to distribute messages, or events, from publishers to subscribers. These negotiator then perform contented based direction-finding of actions at the application-level seeing as content-based direction-finding at the system level.

VI. PERFORMANCE ANALYSIS

A. Performance Metrics for FastReplica

The accede to occasion are denote the relocate occasion of F form file from the unique swelling N_1 to node N_0 are maintained to calculated at node N_i . The two important performance metrics are irregular and utmost all replica of various node are considered in naive setting all the nodes and links are harmonized, and very few node can hold up n network associates to various important server at B bytes/sec. at that time,

$$\text{Time duration}_{\text{partition}} = \text{Size1 (F1)} / (\text{nx}B1)$$

$$\text{Time duration}_{\text{partition}} = \text{Size1 (F1)} / (\text{nx}B1)$$

For FastReplica

$$\begin{aligned} \text{Time1} &= \text{Time1}_{\text{distribution}} + \text{Time1}_{\text{collection}} \\ &= 2 \times \text{Size (F1)} / (\text{nx}B1) \end{aligned}$$

For manifold Unicast

$$\text{Time} = \text{Size (F)} / B$$

VII. CONCLUSION

An method called FastReplica is applied for well-organized and dependable replication of bulky in the middle of the in receipt of nodes to broadcast . The sub files inside the duplication set up are ordered up in the general form. The old document is contented. The durability may be attained by the algorithm by bunch in the even part of duplication groups position by position well-organized i.e. by construction. The presentation results are outlook the the FastReplica in various larger network from side to side Overlay Broker the query were routed to the appropriate SONs and from side to side which the complex transfer load is very much reduced. seeing as the client downloads the each portion of the file from different peers thus the system load is reserved balanced.

REFERENCES

1. Amutharaj, J., & Radhakrishnan, S. (2007, February). The Dominating Set Theory Based Semantic Overlay Networks for Efficient Content Distribution. In Signal Processing, Communications

- and Networking, 2007. ICSCN'07. International Conference on (pp. 228-232). IEEE.
2. J. Byers, J. Considine, M. Mitzenmacher, and S. Rost, "The Informed content delivery across adaptive overlay networks," *Proc. ACM SIGCOMM*, Pittsburgh, PA, Aug. 2004, pp. 47-60.
3. Arturoo Crespo, Hector Garciaa Molina, "Semantic Overlay Networks for P2P Systems," *Google Technologies Inc.*, Stanford University January 2003.
4. Thorstenn Struffee / Dietrich Reschke, "Efficient Content Distribution in semi-decentralized Peer-to-Peer-Networks" *Proceedings of 8th Netties Conference*, Technische Universität Ilmenau, September 30th to October 2nd 2002.
5. Kirrk L. Johnsons, John F. Carr, Markk S. Day, M. Franss Kaashoek, "Certain Measured Performance of Content Distribution Networks" *Technical Report*, SightPath, Inc. 135 Beaver Street, Waltham, MA 02452, USA, 2000.
6. Michael O. "Rabinhood ,The Efficient dispersal of information for security, load balancing, and fault tolerance," *Journal of the ACM (JACM)*, v.36 n.2, p.335-348, April 1989.
7. L. Cherkasova, J. Lee, "FastReplica: Efficient Large File Distribution within Content Delivery Networks", *Proc. of the 4th SENIX Symposium on Internet Technologies*, March, 2002.
8. Pablo Rodriques and Ernst W. Biersack, "Dynamic Parallel Access to Replicated Content in the Internet" , *IEEE / ACM Transactions on Networking*, vol. 10 No.4, Aug 2002.
9. J. Byers, M .Luby, M. Mitzenmacher, A. Rege., "A Digital Fountain approach to reliable distribution of bulk data", *Proc. of ACM SIGCOMM*, 1998
10. Project JXTA: Technical Specification Sun Microsystems, Inc. April 25, 2001. "Survey of Content Delivery Networks (CDNs)", web source.
11. Mohammadi Mallil, Chadii Barakatt, Walid Dabbous, "An Efficient Approach for Content Delivery in Overlay Networks", *Projet Planet*, INRIA-Sophia Antipolis, France, 2005.