International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 4, ISSN No. 2455-2143, Pages 237-240
Published Online February-March 2017 in IJEAST (http://www.ijeast.com)

CONTENT DISTRIBUTION VIA P2P NETWORKS: A SURVEY

Nandhini Vineeth, Ashwin Kandel, Karan Uday, Suhail Muhammed
Department of Computer Science and Engineering
BMSCE College, Bangalore, Karnataka, India

***Abstract*— today the major portion of internet traffic is comprised of content such as videos, images and music files. Distributing such content to the end users efficiently is a challenge which has been addressed by various client-server architecture models such as server farms, web proxies and CDNs (Content distribution System). This paper surveys how an alternative architecture called p2p (peer to peer) can be used in assisting the client server architecture to make the delivery more efficient. This paper presents the state of content distribution in the internet, p2p architectures available, and working of Bit Torrent and Web RTC methods.**

Keywords— **client-server, cdn, p2p, bittorrent, webrtc,**

I. INTRODUCTION

The internet was made to form a network of systems that could connect and communicate with each other seamlessly. A major portion of the internet traffic in its earlier days was occupied by communication services. In spite of the existence of VoIP, chat and other sorts of communication applications, most of the communication traffic revolved around email communication.

But over the last two decades, we have observed a major paradigm shift in the primary role of the Internet from that of enabling communications, to that of distribution of content. This dramatic shift is explained by the transition of offline activities—such as reading news, browsing TV or movie listings, and checking the weather—to online. The websites that serve as online resources have become tremendously popular. An example of this would be that of how streaming services like YouTube and Netflix account for 70% of the internet traffic.

II. CURRENT STATE

Nowadays, almost anything and everything on the internet is served via websites. The Web Designing market has grown tremendously in a short span of a couple of years, owing to the boom of the media and entertainment markets and their need to expand their reach to every inch of the globe. As a result, the web pages being designed today showcase a ton of media content, to be as user friendly, dynamic and creatively intuitive as possible, thereby increasing their viewership. Most of the websites in the internet at present make use of CDNs (Content Delivery Networks) to assist the distribution of their website and contents.

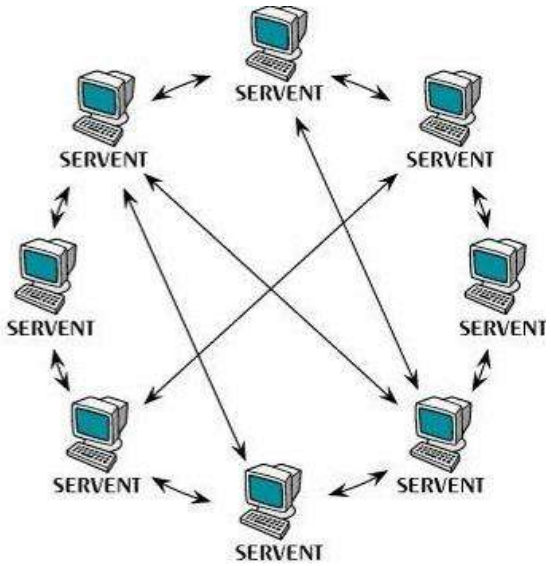
CDN is a group of servers, or data centers, located in several geographical areas to serve end-users. Simply speaking, a CDN helps deliver contents to end-user faster. The "content" can be web, video or application. For example, in the case of the www, end-users send their request to the web server, and the web server redirects this request to the CDN. The CDN then makes a decision to choose which server, out of a vast array of servers spread across several domains connected to it, to use to serve this user. One simple way is to choose a server that's geographically close to the user, but in practice more complicated schemes are used to make sure the content is made available faster, and can be sent over to the user as quickly as possible.

CDN is a client-server architecture. CDN companies such as Akamai are providing the services to web sites such as Facebook, YouTube to serve their content globally. Such systems are efficient but are prone to failure if not handled carefully.

III. PEER TO PEER NETWORKS (P2P)

p2p network is an architecture which is made up of entities called as peers. A peer can act both as server and client so it is also called as servent as shown in figure 1. As a client, it consumes content from other peers. As server, they pool in their resources such as processing power and bandwidth to provide the content to other peers. Such architecture do not require dedicated infrastructure or centralized control and can coordinate, scale and self-organize in the presence of a highly transient population of nodes, network and computer failures.

International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 4, ISSN No. 2455-2143, Pages 237-240
Published Online February-March 2017 in IJEAST (http://www.ijeast.com)



The content distribution in a p2p network is highly scalable as the bandwidth of all participating end nodes are utilized during the distribution process. In fact such networks are inherently self-scalable meaning bandwidth capacity of network increases with increase in participating nodes [2]. Thus p2p networks are inverse scaling. This feature can solve the problem of flash crowding [3] which is a phenomenon where content in a network suddenly becomes popular and there is high demand for it. Furthermore, content distributor can lower the setup and running cost of serving the bandwidth intensive content to thousands of simultaneous users both in the internet and private network[4][5].

Peer to Peer Infrastructures
This contains p2p based infrastructure which are not applications by themselves, but provides peer to peer based services and frameworks. They mainly provide these services:

Routing and location: They provide a system, for routing the request and message efficiently and with fault tolerance. Also they provide proper way for locating the peers and content. Different algorithms such as Chord [6], CAN [7], and Kademlia[8] has been developed for this purpose.

Anonymity: These systems are designed with the purpose of hiding the identity of users and making them anonymous in the network. One of the popular example of such service is Tor which uses onion routing [9] to accomplish the task

IV. BITTORRENT

BitTorrent is a protocol for sharing files between set of peers quickly and easily. It was created in 2001 by Bram Cohen and since then it has contributed to major portion of internet traffic [10]. There are various software like uTorrent, Transmission, Btorrent which act as BitTorrent client allowing peers to share data.

Mechanism
The mechanism for handling the distribution of file using BitTorrent is described in the protocol specification [11]. It can be summarized as:

Describing Content
Instead of distributing entire content as one chunk, the content is divided in several chunks of fixed size. Each chunk is cryptographically hashed using SHA-1 and hash values of all chunks are embedded into a file along with ordering of file. That file is referred to as torrent file. The chunking of file helps in maintaining the integrity of content as well as helps in efficient content distribution

Finding Peers
Peers use tracker server to find the list of other peers who are sharing the same file. The location of tracker server is embedded in torrent file. Tracker server maintains a list of peers for each content which is called as swarm. Therefore a peer must possess the torrent file to download the content from other peers.

Preventing Free-riders
Free-riders are the peers which only use the resource but do not share their resource. Such nodes are harmful to the overall network or peers in p2p system. BitTorrent uses tit-for-tat mechanism to avoid such peers. This algorithm throws away such free-riders from the network which is called as choking.

Study and Analysis
Pouwelse et. al. in their work[12] have analyzed the real BitTorrent system called supernova which has large user base. The study resulted in collection of measurement data of that system utilizing the processing speed of 100 super computers which provided processing power in a distributed way. The paper contains analysis of overall system activity, availability, integrity, flash crowd, and the download performance of supernova. These data provide a model and insight into the behaviour of large p2p systems. The model presented by the study proves the integrity and availability demonstrated by p2p systems.

Qiu et. al. [13] developed simple models to study the performance of Bittorrent. They presented the simple fluid model and studied the scalability, performance, and efficiency of such a file sharing mechanism. Specifically they obtained the expression for average number of seeds, downloaders and downloading time as a function of the arrival times. This fluid model provides insight into network performance. The study concluded that a fluid model can capture the performance of the system if arrival rate is small.

International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 4, ISSN No. 2455-2143, Pages 237-240
Published Online February-March 2017 in IJEAST (http://www.ijeast.com)

V. WEBRTC

With the rise in popularity of content distribution through browsers and mobile applications, came the need for more convenient and standard methods of real-time communication. WebRTC is an API definition drafted by the W3C that provides real time communication capabilities to browsers and mobile phones, also facilitating voice, video and data stream transmission [14]. The API enables users to establish bidirectional communication channels with other peers and compliant servers. Our primary focus in this discussion will be p2p aspect of WebRTC.

Establishing Connection

Though we are moving forward to a more peer based architecture, servers are still required in order to coordinate functions such as finding peers, signaling and NAT traversal. WebRTC enables the use of STUN (Session Traversal Of User Datagram Protocol) & TURN (Traversal Using Relay Net) servers with NAT traversal techniques for establishing a bidirectional channel with another peer. The API needs a mechanism to coordinate communication and to send control messages, a process known as signaling. Signaling protocols are not specified by WebRTC and are left to the discretion of the developer [15].
When we want to establish a connection with a peer we create an RTCPeerConnection object with configurable parameters. Over this connection we can use the RTP media API for media streams or the peer-to-peer data API for sending and receiving arbitrary data streams [16].

Peer-to-Peer Data API

WebRTC specifies an RTCDataChannel API used for bidirectional exchange of generic data. This opens up a world of possibilities from p2p file sharing to multiplayer games. The API supports a flexible set of data types such as String, Blob, ArrayBuffer and it is also designed to mimic WebSockets.

RTCDataChannel can work either in an *unreliable* mode (analogous to User Datagram Protocol or UDP) or a *reliable mode* (analogous to Transmission Control Protocol or TCP).

- Reliable mode guarantees the transmission of messages and also the order in which they are delivered. This results in an overhead, thus potentially making this mode slower.
- Reliable mode guarantees the transmission of messages and also the order in which they are delivered. This results in an overhead, thus potentially making this mode slower.
- Unreliable mode does not guarantee the transmission of every message to the other side or in what order they get there. This removes the overhead, allowing this mode to work much faster.

With such diverse modes, a developer is given ample freedom to build more engaging applications by using high performance, low latency connections via RTCDataChannel.

VI. CONCLUSION

BitTorrent has proven to be one the most successful peer to peer architecture that does the work of distributing file over networks efficiently. It uses networking stack provided by the operating system which cannot be used by applications running in the browser. Now, with the availability of pure peer to peer communication between browsers through WebRTC data channel, various ways of sharing files using it comes into picture. One such architecture is peer CDN architecture which defines protocol similar to BitTorrent on top of WebRTC to distribute web page contents in a p2p fashion.

VII. REFERENCE

[1] Schollmeier, Rüdiger. "A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. Proceedings of "Peer-to-Peer Computing, 2001", IEEE First International Conference on. 2001.

[2] Karagiannis, Thomas, Pablo Rodriguez, and Konstantina Papagiannaki. "Should internet service providers fear peer-assisted content distribution?" Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement. USENIX Association, 2005

[3] Jung, Jaeyeon, Balachander Krishnamurthy, and Michael Rabinovich. "Flash crowds and denial of service attacks: Characterization and implications for CDNs and web sites." Proceedings of the 11th international conference on World Wide Web. ACM, 2002.

[4] Li, Jin. "On peer-to-peer (P2P) content delivery." Peer-to-Peer Networking and Applications 1.1 (2008): 45-63.

[5] Stutzbach, Daniel, Daniel Zappala, and Reza Rejaie. "The scalability of swarming peer-to-peer content delivery."

International Conference on Research in Networking. Springer Berlin Heidelberg, 2005.

[6] Stoica, I., Morris, R., Karger, D., Kaashoek, M. F., & Balakrishnan, H. (2001). Chord: A scalable peer-to-peer lookup service for internet applications. ACM SIGCOMM Computer Communication Review, 31(4), 149-160.

[7] Ratnasamy, S., Francis, P., Handley, M., Karp, R., & Shenker, S. ACM - A scalable content-addressable network, 2001,Vol. 31, No. 4, pp. 161-172. [Use the same format /sequencing for all for uniformity]

[8] Maymounkov, Petar, and David Mazieres. "Kademlia: A peer-to-peer information system based on the xor metric." International Workshop on Peer-to-Peer Systems. Springer Berlin Heidelberg, 2002.

[9] Goldschlag, David, Michael Reed, and Paul Syverson. "Onion routing." Communications of the ACM 42.2 (1999): 39-41.

[10] Schulze, Hendrik, and Klaus Mochalski. "Internet Study 2008/2009." Ipoque Report 37 (2009): 351-362.

[11] BEP_0003. The BitTorrent Protocol Specification, Cohen. http://www.BitTorrent.org/beps/bep_0003.html

[12] Pouwelse, J., Garbacki, P., Epema, D., & Sips, H. (2005, February). The BitTorrent p2p file-sharing system: Measurements and analysis. In International Workshop on Peer-to-Peer Systems (pp. 205-216). Springer Berlin Heidelberg.

[13] Qiu, D., & Srikant, R. (2004, August). Modeling and performance analysis of BitTorrent-like peer-to-peer networks. In ACM SIGCOMM computer communication review (Vol. 34, No. 4, pp. 367-378). ACM.

[14] Sheetal Thakkar, Kinjal Thakkar, Bhavika Bhanushali, Aman Arora, Dishant Chawla;" Research and Review of WebRTC- the Next Generation of Web-based communication".International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 6, June 2015;ISSN: 2277 128X"

[15] Sam Dutton, Getting Started with WebRTC. http://www.lcis.com.tw/paper_store/paper_store/20162260286256T-20163258513437.pdf

[16] Bergkvist, Adam, D. Burnett, and Cullen Jennings. "A. Narayanan," WebRTC 1.0: Real-time Communication between Browsers." World Wide Web Consortium WD WD-webrtc-20120821 (2012).