IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-4, Issue-3 E-ISSN: 2347-2693

A Survey on Content Delivery of Web-Pages

Aaqib Bashir[1] and Prof. T. H. Gurav[2]

[1*,2]*Department of Computer Engineering,*
Smt. Kashibai Navale College of Engineering, India

www.ijcseonline.org

Received: Feb/22/2016 Revised: Mar/06/2016 Accepted: Mar/14/2016 Published: Mar/31/2016

Abstract— Content delivery networks (CDNs) play an important role in today's Internet. They serve a large multimedia data over the Internet and resolve the issues of scalability and network congestion. The content providers are no longer the sole generators of the content in CDN system but user are also involved. User generated content (UGC) is emerging as one of the dominant forms in the multimedia industry and is reshaping the way users watch video, with millions of content producers and consumers. In particular, multimedia sites are creating new viewing patterns and social interactions that influence the way users consume data. There also exists various techniques such as collaborative caching, context-aware recommendations, etc. that can help in efficient content delivery. In this paper, we present a survey on various techniques and methods employed in the delivery of multimedia content. We present a way to optimize the system performance, and characterizing the tradeoff between cost and quality of service in content delivery.

Keywords— ***Content delivery networks, data mining, quality of service, UGC, user behaviour.***

I. INTRODUCTION

As Web-sites become popular, they are vulnerable to flash crowd problem, in which the request load overwhelms the site's infrastructure, such as front-end Web server, network equipment, or bandwidth or the back-end transaction processing infrastructure. This overload can crash the whole site or cause high response time, resulting in loss of revenue or negative customer attitude. More efficient content delivery over the Web has become an important element for improving web performance.

Content Delivery Networks (CDNs) have been proposed to improve accessibility, maximize the bandwidth and maintain correctness through replication of content. With CDNs, content is distributed to cache or surrogate servers located close to the users, resulting in fast, reliable applications and Web services for users. CDNs mainly consist of content providers, CDNs operators and end users.

In today's dynamic Internet landscape, it is more significant for content and service providers to understand the requirements and demands of particular users. So for this reason, content providers are customizing their own content delivery networks to provide content services to their end users. Apart from the content providers, the user are also active participants of content generation and delivery. User Generated Content (UGC) is gaining popularity in today's multimedia Websites such as YouTube, Netflix, etc.

The end users who consume data have certain viewing patterns. We can say that, there exists a pattern how users consume data i.e. there are certain group of users who are interested in certain group of contents. Conventional CDNs do not use the historical data while serving contents to end users. This is not efficient in today's content delivery, where users show a certain preference for content. So, the crowd pattern plays an important role in serving content to particular user groups [1].

Another aspect that affects or influences the content delivery is the social propagation. The contents that are propagated using the social media networks are found to be more popular. Popularity of contents is determined by the sharing and re-sharing of contents on social-networks. With the help of various available algorithms such as collaborative caching, context-aware recommendations, etc. the efficiency of content delivery can be improved, as well as quality of service.

The rest of the paper is organized as follows. Section II presents related work. Section III discusses the proposed system in detail. Finally, Section IV concludes the paper.

II. RELATED WORK

In this section we discuss existing work done by the researchers on content delivery.

M. Cha et al. provided in depth study of YouTube and other similar UGC systems on the basis of large data collection. Author explored the data analysis based on popularity distribution, popularity evolution, and content duplication of user generated contents. Author studied behavior of user and recognized the solution that shape the popularity distribution. They provided the opportunities to influence the latent demand for niche videos that are not reached today

due to information filtering effects or other system scarcity distortions [2].

F. Chen et al. addressed the issue of incorporation of Web server replicas in storage cloud-based CDNs together with building distribution paths among them to reduce the cost incurred on the CDN providers although satisfying QoS requirements for user requests. They constructed the problem as an Integer Program and offered a variety of offline and online greedy heuristics and also they estimated their performance in opposition to the optimal by means of Web trace-based simulations. In real world, latency can vary because of obstruction, network failures, route changes etc., which is not explicitly addressed in this paper [3].

A. Datta et al. proposed an approach for granular, proxy-based caching of dynamic content which merge the benefits of both proxy-based and back end caching. Proposed approach is able reduce the bandwidth and response times in real-world dynamic Web applications. Page level caching solutions must rely on the request URL to identify pages in cache. Proposed dynamic proxy caching technique allows granular, proxy-based caching where both the content and layout can be dynamic [4].

H. Liu et al. studied content multihoming, by initiating the CMO algorithm and the client adaptation algorithm. These two algorithms are proposed to reduce the cost and improve the performance for content multihoming. Algorithm was designed to calculate assignments of content objects to content distribution networks for content publishers. They also demonstrated a novel, lightweight client adaptation algorithm executing at individual content viewers to achieve scalable, fine-grained, fast online adaptation to optimize the quality of experience (QoE) for individual viewers. Content multihoming algorithms decreases publishing cost by up to 40% and client algorithm executing in browsers decreases viewer QoE degradation by 51% [5].

I. Poese et al. proposed a novel system which permits ISPs to determine and utilize path diversity called Provider-aided Distance Information System (PaDIS). Proposed approach utilizes information only accessible to the ISP to rank any client-host pair on the basis of distance information, such as delay, bandwidth or number of hops. Content providers and distributors may also utilize PaDIS to enhance content replication and delivery strategy due to increased access to meta-information [6].

S. Scellato et al. demonstrated how to extract geographic information from social cascades to improve caching of multimedia files in a Content Delivery Network (CDN). Cache replacement policies use information to guarantee that content relevant to a social cascade is kept close to the users who may be interested in it. Performance evaluation shows that they are able to improve cache hits with respect to cache policies without geographic and social information [7].

G. Szabo et al. introduced a method and verification on how the popularity of content can be guessed very soon by measuring the popularity at an early time. On basis of linear correlation author introduced three models to make predictions about future popularity [8].

Z. Wang et al. studied how social network has reshaped the way video contents are generated, distributed and consumed on today's Internet. They conducted a measurement of a real-world online social network system to study the propagation of the social video contents. Authors propose a propagation-based social-aware replication framework using a hybrid edge cloud and peer-assisted architecture, namely PSAR, to serve the social video contents [9].

H. Yin et al. present the design and deployment experiences with LiveSky, a commercially deployed hybrid CDN-P2P live streaming system. CDNs and P2P are two common techniques used for live streaming, each having its advantages and disadvantages. Authors used a hybrid approach where it inherits the best of both the techniques [10].

III. IMPLEMENTATION DETAILS

This section discusses the proposed system in detail.

System Overview

Figure 1 shows the architectural view of the proposed system. The description of the system is as follows:

User is the first entity in the architecture proposed. The user generates a query using certain keywords. The query generated by the user undergoes the processing. After being processed, the query is matched with the present dataset of the content provider. If the content related to the query is present in the dataset, it undergoes through various phases to provide the desired content, related to query, back to the user.

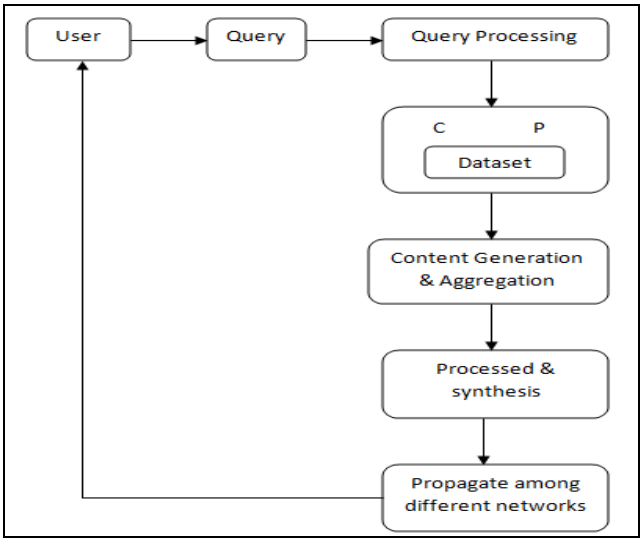

Figure 1: *System Architecture*

Content is generated and aggregated, after the data is found in the dataset. The generated content is put through the processing phases using various algorithms. The algorithm decides the content that will be available to the user, i.e. it provides us with the relevant content. This content is then propagated through various networks, to be delivered back to the user.

The content that has been generated is replicated to various servers to be made available for future use. This content has the probability of being used over and again, based on the crowd pattern and social influence. When this content is replicated to the surrogate servers, the content availability is ensured for future references.

IV. CONCLUSION

This paper presented an all-inclusive survey on content delivery of web-pages. This paper presented an in-depth survey on various methods that can be employed to improve content delivery. In particular, we discussed the context-aware recommendations to schedule the component delivery according to their diverse importance under different contexts, we utilize the users' viewing pattern to replicate contents close to users of potential interests, and furthermore we leverage the social influence to provision resources for social contents that are likely to attract flash crowds.

REFERENCES

[1] Zhi Wang, WenwuZhu,Minghua Chen, Lifeng Sun and Shiqiang Yang,"CPCDN: Content Delivery Powered by Context and User Intelligence", IEEE Transactions on Multimedia, Vol. 17, No. 1, January **2015**.

[2] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, "I tube, you-tube, everybody tubes: Analyzing the World's Largest User Generated Content Video System", in Proc. ACM SIGCOMM Conference on Internet Measurement, pp. **1–14**, **2007.**

[3] F. Chen, K. Guo, J. Lin, and T. La Porta, "Intra-cloud lightning: Building CDNs in the Cloud", in Proc. IEEE International Conference on Computer Communication, pp. **433–441**, Mar. **2012**.

[4] A.Datta, K. Dutta, H. Thomas,D.VanderMeer andK.Ramamritham, "Proxy-based Acceleration of Dynamically Generated Content on the World Wide Web:An approach and Implementation", ACM Transaction onDatabase Systems, vol. 29, no. 2, pp. **403–443**, **2004**.

[5] H. Liu, Y. Wang, Y. Yang, A. Tian and H. Wang, "Optimizing Cost and Performance for Content Multihoming", in Proc. ACM SIGCOMM Conference on Applications, technologies, architectures and protocols for computer communication, pp. **371–382**, **2012**.

[6] I. Poese, B. Frank, B. Ager, G. Smaragdakis and A. Feldmann, "Improving Content Delivery Using Provider-aided Distance Information", in Proc. ACM SIGCOMM Conference on Internet Measurement, pp. **22–34**, **2010**.

[7] S. Scellato, C. Mascolo, M. Musolesi and J. Crowcroft, "Track Globally, Deliver Locally: Improving Content Delivery Networks by Tracking Geographic Social Cascades", in Proc. ACM International Conference on World Wide Web, pp. **457–466**, **2011**.

[8] G. Szabo and B. A. Huberman, "Predicting the Popularity of Online Content", Communications of ACM, vol. 53, no. 8, pp. **80–88, 2010**.

[9] Z. Wang, L. Sun, X. Chen, W. Zhu, J. Liu, M. Chen and S. Yang, "Propagation-Based Social-Aware Replication for Social Video Contents", in Proc. ACM International Conference on Multimedia, pp. **29–38**, **2012**.

[10] H. Yin, X. Liu, T. Zhan, V. Sekar, F. Qiu, C. Lin, H. Zhang and B. Li, "Design and Deployment of a Hybrid CDN-P2P System for Live Video Streaming: Experiences with Livesky", in Proc. ACM International Conference on Multimedia, pp. **25–34**, **2009**.

AUTHORS PROFILE

Mr. Aaqib Bashir is currently pursuing M.E from Department of Computer Engineering, SKN College of Engineering, Pune, India. Savitribai Phule Pune University, Pune, Maharashtra, India -411007. He received his B.Tech degree in Computer Science and Engineering from Islamic University of Science and Technology, Jammu and Kashmir, India. His areas of interest are web and data mining, data analytics, databases.

Prof. T. H. Gurav received her M. E (Computer Networks) Degree from Sinhgad College of Engineering, Vadgaon, Pune, India. She received her B.E (Computer) Degree from Sanjivani College of Engineering, Kopargaon, Maharashtra, India.She is currently working as M.E coordinator and Asst. Prof. (Computer) at Department of Computer Engineering, SKN College of Engineering, affiliated to Savitribai Phule Pune University, Pune, India 411007. Her areas of interest are databases, computer graphics, computer networks.