

Revisiting Caching in Content Delivery Networks

M. Zubair Shafiq[†], Alex X. Liu[†], Amir R. Khakpour[‡]

[†]Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, USA

[‡]Verizon/EdgeCast, Santa Monica, CA, USA

{shafiqmu,alexliu}@cse.msu.edu, amir@edgecast.com

ABSTRACT

Content Delivery Networks (CDNs) differ from other caching systems in terms of both workload characteristics and performance metrics. However, there has been little prior work on large-scale measurement and characterization of content requests and caching performance in CDNs. For workload characteristics, CDNs deal with extremely large content volume, high content diversity, and strong temporal dynamics. For performance metrics, other than hit ratio, CDNs also need to minimize the disk operations and the volume of traffic from origin servers. In this paper, we conduct a large-scale measurement study to characterize the content request patterns using real-world data from a commercial CDN provider.

Categories and Subject Descriptors

C.2.3 [Computer System Organization]: Computer Communication Networks—*Network Operations*

Keywords

Caching; Content Delivery Network (CDN)

1. INTRODUCTION

Most major content publishers such as Netflix, Hulu, Facebook, Pinterest, and CNN use Content Delivery Networks (CDNs), often run by third parties, to bring content closer to their end users. More than 70% of all Internet traffic nowadays is carried by CDNs [6]. A CDN provider typically owns or rents spaces from multiple data centers, called Points of Presence (PoPs), which are geographically distributed around the world. When a content publisher contracts a CDN to host their content, a user request to their content (such as a web page, an image, or a video file) is redirected to the closest PoP via DNS redirection, HTTP redirection, URL rewriting, or anycast [7]. If that PoP does not have the content, then the PoP fetches it from the origin server of the content publisher, sends it to the user, and also caches the

content in storage such as Solid State Drives (SSDs) or Hard Disk Drives (HDDs). Thus, for later requests to the same content, the PoP directly sends its cached copy to the users if the Time-To-Live (TTL) value has not expired. CDNs provide many benefits such as faster content delivery to end users, higher service availability, better user experience, less traffic to content publisher servers, and better protection of content publishers from Denial of Service (DoS) due to their larger capacity of handling legitimate flash crowds and malicious attack traffic.

There has been little prior work on characterization of content requests and caching performance in CDNs. This is partially because CDNs have emerged only for about ten years and have become prevalent only recently as video, image, and other User Generated Content (UGC) has become dominant sources of traffic on the Internet and way passed the capacity that content publishers can handle by themselves. In particular, no caching schemes have been designed particularly for CDNs in prior literature, to the best of our knowledge. CDN providers still commonly use the standard cache replacement algorithms such as Least Recently Used (LRU), Least Frequently Used (LFU), and their minor variants. However, CDNs deserve specialized caching schemes because they differ from other caching systems in terms of both *workload characteristics* and *performance metrics*.

The first step towards designing effective caching schemes tailored for CDNs is to understand the characteristics of content requests because their dynamics strongly impact CDN caching performance. However, little prior work has been done towards this end. Some passive measurement studies on CDNs have been done [3, 10]; however, they are limited to analyzing traffic from CoralCDN, a free and open CDN based on the PlanetLab testbed, whose load is orders of magnitude smaller than that of large commercial CDNs. Some active measurement studies on CDNs have been done [1, 2, 4, 5, 8, 9]; however, they are small-scale studies due to the active nature and they do not provide a characterization of content requests other than those they injected. To the best of our knowledge, a large-scale measurement and characterization of content requests is missing in prior literature.

In this paper, we conduct a comprehensive measurement study to characterize the content requests in large CDNs using real-world data from one of the world's largest commercial CDN providers. Our data set contains detailed traffic logs from multiple PoPs of the CDN provider over the duration of one week in 2012 consisting of more than 600 million requests. All personally identifiable information in our data

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

SIGMETRICS'14, June 16–20, 2014, Austin, Texas, USA.

ACM 978-1-4503-2789-3/14/06.

<http://dx.doi.org/10.1145/2591971.2592021>.

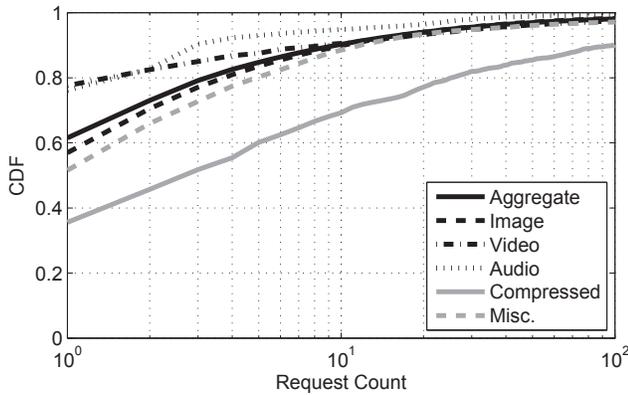


Figure 1: Request count distributions

set (e.g., IP addresses) has been anonymized to protect the privacy of end users without affecting the usefulness of our results. Each record in our logs contains the content publisher’s identifier, hashed URL, object file type, object size in bytes, and the timestamp when the request was received. We analyze content requests from the perspective of their popularity and temporal dynamics.

2. RESULTS

2.1 Object Popularity

We investigate individual object popularity, which is quantified in terms of request count. Figure 1 plots the CDFs of request count for different content types. We observe the expected long-tail distributions for different content types. The distributions highlight that a small fraction of objects appear in most requests and most objects are requested very infrequently. Object popularity distributions have long tails. For example, more than 60% objects are requested only once. We also observe significant variations across different content types.

2.2 Temporal Dynamics

To understand how content requests change over time, we first analyze the timeseries of request counts. Figure 2 shows the timeseries of request counts for different content types. We observe that all timeseries exhibit diurnal variations, with minor variations across different content types. Overall, the request volume peaks from morning till late evening and bottoms out after mid-night. The timeseries of **audio** and **compressed** content are spiky because of their lower volumes. The request volume of **video** and **compressed** content substantially increases on Friday, while other content types exhibit regular diurnal patterns.

3. CONCLUSION

Our measurements characterizing CDN content request patterns highlight the following key findings.

- First, *the distribution of object popularity is highly skewed*; that is, only a small fraction of objects seen by the CDN are requested very frequently while the majority of objects are requested very infrequently. This

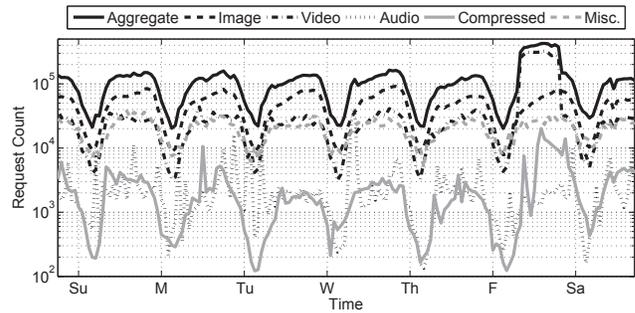


Figure 2: Timeseries of object request volume

implies that CDNs can avoid caching infrequently requested objects so that frequently requested objects can be better served.

- Seconds, *object popularity distribution and temporal dynamics vary significantly for different content types*. This implies that we can partition the total CDN cache space into sub-caches of different sizes where we use different sub-caches to cache objects of different types and this partition dynamically evolves over time.

We plan to investigate these ideas and their implication on content caching in our future work.

4. REFERENCES

- [1] V. K. Adhikari, Y. Guo, F. Hao, V. Hilt, and Z.-L. Zhang. A tale of 3 CDNs: An active measurement study of Hulu and its CDNs. In *IEEE Global Internet Symposium*, 2012.
- [2] V. K. Adhikari, Y. Guo, F. Hao, M. Varvello, M. Steiner, V. Hilt, and Z.-L. Zhang. Unreeling Netflix: Understanding and improving multi-CDN movie delivery. In *IEEE INFOCOM*, 2012.
- [3] M. J. Freedman. Experiences with CoralCDN: A five-year operational view. In *USENIX/ACM NSDI*, 2010.
- [4] C. Huang, A. Wang, J. Li, and K. W. Ross. Measuring and evaluating large-scale CDNs: (paper withdrawn at Microsoft’s request). In *ACM IMC*, 2008.
- [5] B. Krishnamurthy, C. Wills, and Y. Zhang. On the use and performance of content distribution networks. In *ACM SIGCOMM Workshop on Internet Measurement*, 2001.
- [6] C. Labovitz. First Data on Changing Netflix and CDN Market Share, June 2012. <http://www.deepfield.net/2012/06/first-data-on-changing-netflix-and-cdn-market-share/>.
- [7] M. Pathan and R. Buyya. *Content Delivery Networks, Chapter 2: A Taxonomy of CDNs*. Springer, 2008.
- [8] A.-J. Su and A. Kuzmanovic. Thinning Akamai. In *ACM IMC*, 2008.
- [9] S. Triukose, Z. Wen, and M. Rabinovich. Measuring a commercial content delivery network. In *ACM WWW*, 2011.
- [10] P. Wendell and M. J. Freedman. Going viral: Flash crowds in an open CDN. In *ACM IMC*, 2011.