

Broadening the Audience: Popularity Dynamics and Scalable Content Delivery Techniques

Niklas Carlsson

Department of Computer and Information Science
Linköping University, 581 83 Linköping, Sweden

Abstract. The Internet is playing an increasingly important role in today's society and people are beginning to expect instantaneous access to information and content wherever they are. As content delivery is consuming a majority of the Internet bandwidth and its share of bandwidth is increasing by the hour, we need scalable and efficient techniques that can support these user demands and efficiently deliver the content to the users. When designing such techniques it is important to note that not all content is the same or will reach the same popularity. Scalable techniques must handle an increasingly diverse catalogue of contents, both with regards to diversity of content (as services are becoming increasingly personalized, for example) and with regards to their individual popularity. The importance of understanding content popularity dynamics is further motivated by popular contents' widespread impact on opinions, thoughts, and cultures. This article will briefly discuss some of our recent work on capturing content popularity dynamics and designing scalable content delivery techniques.

1 Introduction

When first asked to write an article for this book, I thought of potential articles that I would predict that I would like to read on my (future) 60th birthday and quickly realized that my choice likely would be different than what my dear colleague's choice would be. It is of course the same with content delivered over the Internet. Each user has their individual preferences, and as we are all unique, the range of videos that a set of users select to access will be highly diverse. However, interestingly, some videos will emerge from the masses as much more popular than other videos. In fact, it has been observed that content popularity often follows a heavy-tailed distribution (e.g., [18, 24, 6, 23, 29, 27]). With a heavy-tailed distribution there will be some smaller number of contents that obtain most of the views, while the rest of the contents observe much fewer views.¹

Now, before taking the journey through the content landscape and determine what makes some content more popular than others, we may also want to establish my own

¹ The same holds true for many other things around us [16], including the number of citations associated with the articles that we publish. In the case of my former supervisor, the paper *Adaptive load sharing in homogeneous distributed systems* [21] has collected more citations than his other papers and will likely be the winner at his 60th birthday. For Professor Shahmehri it looks to be a race down to the wire as she has three articles [15, 1, 3] within less than 10 citations away from each other, according to Google scholar.

reading preferences and how it may differ from yours. As numerous top researchers have published their top-10 lists of influential articles they have found particularly insightful and recommend others to read (see for example a very nice series published in ACM Computer Communication Review a few years back), I instead briefly mention two inspiration sources that I can identify with (and which I would recommend somebody that would celebrate their 60th birthday reading). First, on the list is *PhD Comics* [14]. This book series (also delivered on the Internet) gives a unique and valuable perspective into the life of graduate students. It is definitely a must read for any graduate student, but I also think that it can work as a good tool for senior faculty to remember the perspective of the students. Second, as I always have enjoyed playing ice hockey and am doing research within the area of performance analysis of networks and distributed systems, I find the article *The top ten similarities between playing hockey and building a better Internet* [2] both entertaining and insightful.

So, having established that our reading preferences/interests and movies choices likely are different(!), let us consider what makes one video more popular than another and how we best server a set of contents. For this purpose, we will first discuss some of our work on understanding content popularity dynamics (Section 2) and then discuss some of our works on scalable content delivery (Section 3).²

2 Popularity dynamics

Workload characterization and modeling are important tools when building an understanding of system dynamics and identifying design improvements. In this work we have measured, characterized, and modeled longitudinal workloads, including:

- Two simultaneously collected 36-week long traces of the weekly YouTube file popularity of over one million videos [5].
- Two simultaneously collected 48-week long traces capturing the popularity dynamics observed locally at a university campus and globally across seven hundred unique BitTorrent trackers and 11.2 million torrents (contents) [8].

Our prior work on file sharing popularity [29, 18] had shown that the file popularity across shorter time intervals are more Zipf-like than file popularity distributions based on longer time periods, including the life-time views displayed at various content sharing sites, such as YouTube. The above traces allow us to take a closer look at the popularity churn and how it evolves over time.

Using our longitudinal YouTube traces, we develop and evaluate a framework for studying the popularity dynamics of user-generated videos [5]. We present a characterization of the popularity dynamics, and propose a model that captures the key properties of these dynamics. While the relative popularities of the videos within our dataset are highly non-stationary, we find that a simple model that maintains statistics about three sets of videos can be used to accurately capture the popularity dynamics of collections of recently-uploaded videos as they age, including key measures such as hot set churn statistics, and the evolution of the viewing rate and total views distributions over time.

² For an up-to-date list of publications on these (and other) topics, as well as links to our public dataset, please see <http://www.ida.liu.se/~nikca/>.

Using the BitTorrent traces, we compare the popularity dynamics as seen on a university campus compared to globally [8]. While the paper provides insights that may improve the efficiency of content sharing locally, and thus increase the scalability of the global systems, for the purpose of this article, one observation that may be particularly interesting is that we found that campus users are early adapters (in the sense that they typically download files well before the time at which the global popularity of the files peak) for all new content except for music files, for which the campus users are late to download. As Professor Shahmehri has spent a considerable amount of time in the academic environment, this may suggest that an up-to-date music collection may not be an appropriate birthday present? Determining if this is the case is left for future work!

Another noteworthy contribution towards better understanding video sharing popularity involved the use of clones [4]. Well, not the controversial, genetically created kind, but the kind that you may find on video sharing sites. This study set out to understand the *content-agnostic* factors that impact video popularity. While some popularity differences arise because of differences in video content, it is more difficult to accurately study content-agnostic factors. For example, videos uploaded by users with large social networks may tend to be more popular because they tend to have more interesting content, not because social network size has a substantial direct impact on popularity. Leveraging the use of video clones, we developed and applied a methodology that allows us to accurately assess, both qualitatively and quantitatively, the impacts of content-agnostic factors on video popularity. Interesting findings include the observation that a strong linear *rich-get-richer* behavior is observed when controlling for video content. We also analyze a number of phenomena that may contribute to rich-get-richer, including the first-mover advantage and search bias towards popular videos.

3 Scalable content delivery

Today, content delivery applications consume a majority of the Internet bandwidth. With continued rapid growth in demand for such applications anticipated, the problem of cost-efficient and/or sustainable content delivery becomes increasingly important. For efficient delivery, protocols and architectures must scale well with the request loads (such that the marginal delivery costs reduce with increasing demands). Using scalable techniques can allow a content distributor to handle higher demands more efficiently, and/or to offer its existing customers better service while reducing its resource requirements and/or delivery costs.

A variety of techniques have been studied to improve the scalability and efficiency of content delivery, including *replication* [20, 30], *service aggregation* [7, 25, 22, 13], and *peer-to-peer* [17] techniques. With replication, multiple servers (possibly geographically distributed as in a CDN) share the load of processing client requests and may enable delivery from a nearby server. With aggregation, multiple client requests are served together in a manner that is more efficient than individual service. Finally, with peer-to-peer techniques, clients may contribute to the total service capacity of the system by providing service to other clients.

While much work has considered various scalable solutions, there is a lack of literature considering the problem of cost-efficient content delivery, in which the application

incurs both a network delivery cost (e.g., from cross ISP traffic or, more generally, operation/energy costs at Internet routers) and costs at the servers (e.g., due to cost of ownership, energy, or disk bandwidth). Using a batched service model for video-on-demand [11] and a digital fountain model [10] we determine optimal server selection policies for such an architecture, and derive analytic expressions for their associated delivery costs. Our framework also allows us to compare classes of server selection policies and their optimal representatives. We conclude that server selection policies using dynamic system state information can potentially yield large improvements in performance, while deferred rather than at-arrival server selection has the potential to yield further substantial performance improvements for some regions of the parameter space [11]. We also argue that an architecture with distributed servers, each using digital fountain delivery, may be an attractive candidate architecture when considering the total content delivery cost [10]. The importance of these contributions may be further augmented by the potential of increasing energy costs and carbon taxes.

Another content delivery aspect that often is forgotten is how to best serve the entire catalogue of contents available to a content provider. Often the focus is on how to best serve the most popular contents. However, in practice there is a long tail of less popular contents which have a high aggregate request rate (even though each content item on its own has a small request rate) [18, 29]. In recent works, we have presented solutions that address the complete catalogue of files.

Recently we presented and introduced the idea of torrent inflation (dynamic bundling) [12]. In contrast to static bundling (a pre-determined file collection grouped together by the publisher) [28], with dynamic bundling, peers may be assigned complementary content (files or parts of files) to download at the time they decide to download a particular file. This additional flexibility allow dynamic bundling to adapt to current popularities and peers downloading the same popular file to help in the distribution of different less popular files. As we observed in Section 2, this is important as file popularities typically are highly skewed and the download performance of small swarms is poor [19]. In general, our dynamic bundling approach has been found to improve download times and improve file availability for lukewarm (niche) contents. We have also present a prototype implementation [31] and evaluated dynamic peer-based policies with the aid of stochastic games and a Markov Decision Process (MDP) [26].

Thus far, this section has discussed server-based and peer-assisted solutions to the scalability problem. However, recently cloud-based solutions have become a cost-effective means of on-demand content delivery. To understand the tradeoffs between various system design choices, including server bandwidth, cloud and peer resources, we have analyzed a hybrid peer-assisted content delivery system that aims to provide guaranteed average download rate to its customers [9]. Our model and analysis provide valuable insight to the most cost efficient way to deliver a catalogue of content with varying popularity. Among other things, we show that bandwidth demand peaks for contents with moderate popularity, and identify these contents as candidates for cloud-based service. Both bundling and peer-seeding can be used to reduce the need to push content to the cloud, with hybrid policies that combine bundling and peer seeding often reducing the delivery costs by 20% relative to only using seeding.

4 Discussion and future directions

In this paper we have presented some of our recent works on characterizing and modeling the content popularity dynamics, as well as some of our work on scalable and efficient content delivery techniques. Ongoing efforts include the design and evaluation of hybrid solutions that take the environmental footprint into account.

In addition to the impact that popularity dynamics may have on system design, it is also important to note that video dissemination through sites such as YouTube also have widespread impacts on opinions, thoughts, and cultures. As not all videos will reach the same popularity and have the same impact, there is therefore a widespread interest in what makes some videos go viral and reach millions of people while other seemingly similar videos only are viewed by few. While our clone paper gives some insights [4], understanding the relationships between other external and social factors, which impact the video popularity, therefore present interesting future research avenues.

Acknowledgements

Most of this research discussed in this article was done in collaboration with colleagues, including (in alphabetic order): Sebastien Ardon, Martin Arlitt, Youmna Borghol, György Dan, Derek Eager, Nissan Lev-tov, Zongpeng Li, Aniket Mahanti, Anirban Mahanti, Siddharth Mitra, Carey Williamson, and Song Zhang. This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, and CENIIT at Linköping University.

References

1. J. Aberg and N. Shahmehri. The role of human Web assistants in e-commerce: An analysis and a usability study. *Internet Research*, 10(2):114–125, 2000.
2. M. Arlitt. The top ten similarities between playing hockey and building a better internet. *ACM SIGCOMM Computer Communications Review*, 42(2):99–102, Mar. 2012.
3. P. A. Bonatti, C. Duma, N. E. Fuchs, W. Nejdl, D. Olmedilla, J. Peer, and N. Shahmehri. Semantic web policies - a discussion of requirements and research issues. In *Proc. ESWC*, 2006.
4. Y. Borghol, S. Ardon, N. Carlsson, D. Eager, and A. Mahanti. The untold story of the clones: Content-agnostic factors that impact youtube video popularity. In *Proc. ACM SIGKDD*, Beijing, China, Aug. 2012.
5. Y. Borghol, S. Mitra, S. Ardon, N. Carlsson, D. Eager, and A. Mahanti. Characterizing and modeling popularity of user-generated videos. In *Proc. IFIP PERFORMANCE*, Amsterdam, Netherlands, Oct. 2011.
6. L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web caching and Zipf-like distributions: Evidence and implications. In *Proc. IEEE INFOCOM*, Mar 1999.
7. J. Byers, M. Luby, M. Mitzenmacher, and A. Rege. A digital fountain approach to reliable distribution of bulk data. In *Proc. ACM SIGCOMM*, Vancouver, Canada, Sept. 1998.
8. N. Carlsson, G. Dan, M. Arlitt, and A. Mahanti. A longitudinal characterization of local and global bittorrent workload dynamics. In *Proc. PAM*, Vienna, Austria, Mar. 2012.
9. N. Carlsson, G. Dan, D. Eager, and A. Mahanti. Tradeoffs in cloud and peer-assisted content delivery systems. In *Proc. IEEE P2P*, Tarragona, Spain, Sept. 2012.

10. N. Carlsson and D. Eager. Content delivery using replicated digital fountains. In *Proc. IEEE/ACM MASCOTS*, Miami Beach, FL, Aug. 2010.
11. N. Carlsson and D. L. Eager. Server selection in large-scale video-on-demand systems. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 6(1):1:1–1:26, Feb. 2010.
12. N. Carlsson, D. L. Eager, and A. Mahanti. Using torrent inflation to efficiently serve the long tail in peer-assisted content delivery systems. In *Proc. IFIP/TC6 Networking*, Chennai, India, May 2010.
13. N. Carlsson, D. L. Eager, and M. K. Vernon. Multicast protocols for scalable on-demand download. *Performance Evaluation*, 63(8/9), Oct. 2006.
14. J. Cham. *Piled Higher and Deeper (books 1-4)*. Piled Higher and Deeper Pub, 2012.
15. L. Chisalita and N. Shahmehri. A peer-to-peer approach to vehicular communication for the support of traffic safety applications. In *Proc. IEEE International Conference on Intelligent Transportation Systems*, Sept. 2002.
16. A. Clauset, C. Shalizi, and M. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, Nov. 2009.
17. B. Cohen. Incentives build robustness in bittorrent. In *Proc. Workshop on Economics of Peer-to-Peer Systems*, Berkeley, CA, June 2003.
18. G. Dan and N. Carlsson. Power-law revisited: A large scale measurement study of p2p content popularity. In *Proc. IPTPS*, San Jose, CA, Apr. 2010.
19. G. Dan and N. Carlsson. Centralized and distributed protocols for tracker-based dynamic swarm management. *IEEE/ACM Transactions on Networking (ToN)*, to appear.
20. J. Dilley, B. Maggs, J. Parikh, H. Prokop, R. Sitaraman, and B. Weihl. Globally distributed content delivery. *IEEE Internet Computing*, 6(5), Sept/Oct. 2002.
21. D. L. Eager, E. D. Lazowska, and J. Zahorjan. Adaptive load sharing in homogeneous distributed systems. *IEEE Transactions on Software Engineering*, 12(5):662–675, 1986.
22. D. L. Eager, M. K. Vernon, and J. Zahorjan. Minimizing bandwidth requirements for on-demand data delivery. *IEEE Transactions on Knowledge and Data Engineering*, 13(5):742–757, Sept/Oct. 2001.
23. P. Gill, M. Arlitt, Z. Li, and A. Mahanti. Youtube traffic characterization: A view from the edge. In *Proc. IMC*, Oct. 2007.
24. K. Gummadi, R. Dunn, S. Saroiu, S. Gribble, H. Levy, and J. Zahorjan. Measurement, modeling and analysis of a peer-to-peer file-sharing workload. In *Proc. ACM SOSP*, Oct. 2003.
25. A. Hu. Video-on-demand broadcasting protocols: A comprehensive study. In *Proc. IEEE INFOCOM*, Apr. 2001.
26. N. Lev-tov, N. Carlsson, Z. Li, C. Williamson, and S. Zhang. Dynamic file-selection policies for bundling in bittorrent-like systems. In *Proc. IEEE IWQoS*, June Beijing, China.
27. A. Mahanti, N. Carlsson, A. Mahanti, M. Arlitt, , and C. Williamson. A tale of the tails: Power-laws in internet measurements. *IEEE Network*, to appear.
28. D. Menasche, A. Rocha, B. Li, D. Towsley, and A. Venkataramani. Content availability and bundling in swarming systems. In *Proc. ACM CoNEXT*, Dec. 2009.
29. S. Mitra, M. Agrawal, A. Yadav, N. Carlsson, D. Eager, and A. Mahanti. Characterizing web-based video sharing workloads. *ACM Transactions on the Web (TWEB)*, 5(2):8:1–8:27, May 2011.
30. S. Triukose, Z. Wen, and M. Rabinovich. Measuring a commercial content delivery network. In *Proc. WWW*, Mar/Apr. 2011.
31. S. Zhang, N. Carlsson, D. Eager, Z. Li, and A. Mahanti. Dynamic file bundling for large-scale content distribution. In *Proc. IEEE LCN*, Clearwater, FL, Oct. 2012.