# Minimizing the Cask Effect of
# Multi-source Content Delivery

Xi Chen
*Tsinghua University, China*

Zhenhua Li
*Tsinghua University, China*

Zhenyu Li
*ICT, CAS, China*

Tianyin Xu
*UIUC and Facebook, USA*

Ennan Zhai
*Yale University, USA*

Yao Liu
*SUNY Binghamton, USA*

Minghao Zhao
*Tsinghua University, China*

Yunhao Liu
*Tsinghua University, China*

*Abstract*—This paper reveals the performance anomaly (*i.e.,* the decline of delivery speed) when the client upgrades a task from single-source content delivery to multi-source content delivery. This anomaly is mainly caused by two aspects: (1) data sources with different types vary greatly in terms of acceleration reward (AR), and data sources with certain types are particularly easy to become inferior; (2) When the data sources remain fixed for a period of time, the large diversity of participant time (DPT) of data sources disturb the acceleration and the data sources with less participant time are inferior. Combing these insights, we figure out that the multi-source content delivery is limited by the so-called *cask effect*, *i.e.,* the acceleration effect mainly depends on the inferior data sources.

## I. INTRODUCTION

As one of the most fundamental and pervasive Internet services, content delivery has undergone several generations of enabling technologies, including traditional client-server (C/S) models, content delivery networks (CDNs), peer-to-peer (P2P) networks, and cloud-based techniques [1], [2]. Besides, most content delivery industrial systems (*e.g.,* Thunder and QQ Xuanfeng) provide *downloading acceleration function* which is mainly based on *multi-source content delivery* to improve download performance [3], [4]. This means that these systems should integrate multiple content delivery technologies. Once enabling acceleration function, a user client simultaneously fetches different parts of the requested content from multiple data sources through various content delivery techniques and protocols. Seen in a broader context, this function can be applied in internet video. The efficacy is indeed confirmed by previous studies [5], [6].

However, these prior efforts fall short of providing stable and high QoE as the interactions between a user client and the multiple data sources become far more complex today than those were a few years ago. Specifically, we observe the large dataset from M-Downloader, a multi-source content delivery component operated by the leading commercial Internet provider Tencent in China, and evaluate its core function, *i.e.,* accelerating download by upgrading the initial single-source content delivery to multi-source content delivery. Disappointingly, we find that this function sometimes fails (about 23%) to meet the goal of improving delivery speed,

which makes the delivering quality of experience (QoE) far from satisfactory [7]–[9].

This paper conducts the first empirical study to quantitatively understand the anomaly of multi-source content delivery, including its root causes for the design of relevant systems. Our study is based on M-Downloader. It serves 179M (million) file download requests issued from 37M users (including both PC and mobile clients) on a daily basis. For each request, M-Downloader can use up to seven types of data sources, including original C/S links (mostly in HTTP and FTP), free C/S mirrors, *charged* C/S mirrors, free CDNs, *charged* CDNs, ISP caches, and P2P data swarms (mostly in BitTorrent and eMule). Here charged means that M-Downloader must pay for the upload traffic of the used data sources.

Our analysis reveals two key insights. First, data sources with different types are highly heterogeneous in terms of download performance, thus data sources with certain types are particularly easy to become the weak point during multi-source content delivery and only have little acceleration reward (AR). Second, when the data sources remain fixed for a period of time, the core factor that affects the acceleration is the diversity of participation time (DPT, refer to Equation (1)) of data sources. The larger the DPT, the stronger interference the multi-source content delivery, the poorer the acceleration effect, and even an abnormal deceleration may occur. Combing these two insights, we figure out that the multi-source content delivery is subject to *cask effect* to some extent, which could be indicated by the ratio DPT/AR. This cask effect can provides solid experiences and helpful heuristics to the designers of relevant systems.

**Roadmap.** The reminder of the paper is organized as follows. § II describes the dataset from M-Downloader. § III presents the in-depth analysis of multi-source content delivery. Finally, we conclude the paper in § IV.

## II. DATASET

To understand the performance characteristics of multi-source content delivery, we study a large-scale dataset collected from the M-Downloader system. The dataset contains the complete running logs of the system during a whole week (July 13–19, 2015), involving 1,364,122,406 download
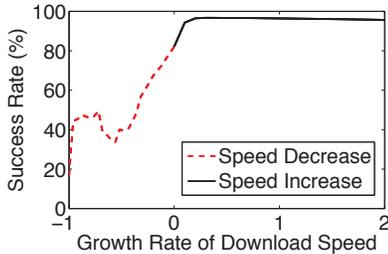
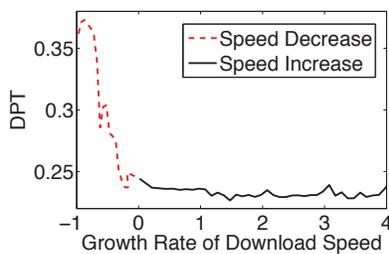Fig. 1. CDF of growth rate of download speed.



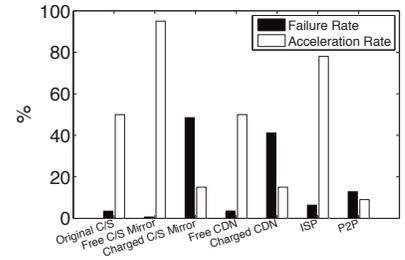Fig. 2. Diversity of participation time *vs.* growth rate of download speed.



Fig. 3. Accelerate rate and download failure rate of different types.

tasks, 57,538,801 users, and 9,827,109 unique files. Among these download tasks, the majority (59%) utilized multiple ($\geq 2$) data sources, and the remainder (41%) only used the original (single) data source. From the dataset, we find that M-Downloader can use up to seven types of data sources for multi-source content delivery. The seven types of data sources cover almost all popular content delivery techniques and protocols at present, and they are original C/S , free C/S mirrors , charged C/S mirrors, charged CDN, Free CDN, ISP caches and P2P swarms respectively.

## III. PERFORMANCE ANALYSIS

To understand the performance characteristics of multi-source content delivery, we study a large-scale dataset collected from the M-Downloader system and evaluate the performance of M-Downloader via comprehensive real-world measurements.

We firstly reveals an unexpected performance degradations appeared in the download speed when the download task is upgraded from original single-source content delivery to multi-source content delivery. Figure 1 shows that 23% of downloads (depicted as the red dashed curve) become slower after being upgrading. Additionally, approximately 37% of downloads are trivially accelerated by almost zero KBps.

Through data-driven analysis, we examine the correlation between the diversity of participation time (DPT) and the growth rate of download speed. Here, DPT is measured by the standard deviation divided by the range, *i.e.,*

$$\frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (T_i - \overline{T})^2}}{T_{max} - T_{min}}, \quad (1)$$

where $T_i$ denotes the participation time of the $i$-th data source, and $\overline{T}$ denotes the average participation time of the data sources used. Figure 2 quantifies the obviously negative correlation between the diversity of participation time and the growth rate of download speed. Specifically, when the diversity of participation time is small ($< 0.2$), a download task can usually benefit from using multiple data sources.

Furthermore, we quantitatively compare the data sources of different types. Figure 3 shows that among seven types of data source, free C/S mirror has the best delivery quality, followed by ISP, free CDN and original C/S, with charged

CDN, charged C/S mirror and P2P worst, considering both download failure rate and acceleration rate. This indicates that the type of data source is the fundamental attribute of influencing acceleration result and even download result.

## IV. CONCLUSION

In this paper, we reveal a performance anomaly of multi-source content delivery by analyzing a large-scale dataset provided by the M-Downloader system. We investigate its root causes for addressing the performance anomaly. Finally, we conclude that multi-source content delivery subjects to the so-called *cask effect*, *i.e.,* the weak points can disturb the acceleration effect and download result.

## REFERENCES

[1] Y. Huang, T. Z. Fu, D.-M. Chiu, J. Lui, and C. Huang, "Challenges, design and analysis of a large-scale P2P-VoD system," in *Proc. of ACM SIGCOMM*, 2008.

[2] C. Wu, B. Li, and S. Zhao, "On dynamic server provisioning in multichannel P2P live streaming," *IEEE/ACM Transactions on Networking (TON)*, vol. 19, no. 5, pp. 1317–1330, 2011.

[3] P. Dhungel, K. W. Ross, M. Steiner, Y. Tian, and X. Hei, "Xunlei: Peer-assisted download acceleration on a massive scale," in *Proc. of PAM*, 2012, pp. 231–241.

[4] Z. Li, Y. Huang, G. Liu, F. Wang, Y. Liu, Z.-L. Zhang, and Y. Dai, "Challenges, designs, and performances of large-scale open-p2sp content distribution," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 11, pp. 2181–2191, 2013.

[5] 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. of ACM Multimedia*, 2009.

[6] P. Aditya, M. Zhao, Y. Lin, A. Haeberlen, P. Druschel, B. Maggs, and B. Wishon, "Reliable client accounting for P2P-infrastructure hybrids," in *Proc. of USENIX NSDI*, 2012.

[7] S. Sundaresan, W. De Donato, N. Feamster, R. Teixeira, S. Crawford, and A. Pescapè, "Broadband Internet performance: a view from the gateway," in *Proc. of ACM SIGCOMM*, 2011.

[8] Z. L. Chen, Guihai, *Peer-to-Peer network: Structure, application and design*. Tsinghua University Press, 2007.

[9] Z. Li, Y. Huang, Y. Huang, Z. L. Zhang, and Y. Dai, "Maximizing the bandwidth multiplier effect for hybrid cloud-p2p content distribution," in *IEEE IWQos*, 2012, pp. 1–9.