Path-Aware Peer-Assisted Web Content Delivery against Network Failures

Wen Hu[1], Zhi Wang[2], and Lifeng Sun[1]

[1]Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science and Technology, Tsinghua University
[2]Graduate School at Shenzhen, Tsinghua University
{hu-w12@mails., wangzhi@sz., sunlf@}tsinghua.edu.cn

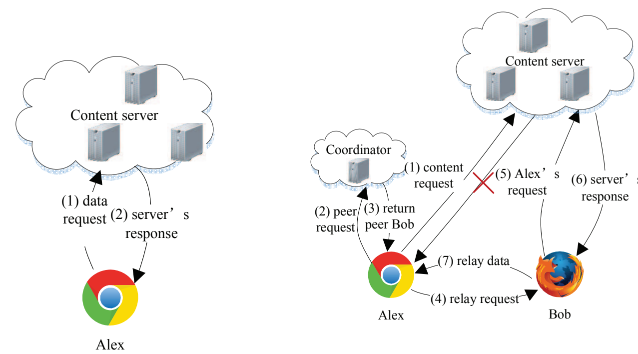
***Abstract*—Popularly used to distribute a variety of multimedia contents in today's Internet, HTTP-based web content delivery still suffers from failures occurring both in the network and at the servers. Limited by the deployment of edge servers, it is hard for a CDN to successfully deliver contents to all users under these network failures. Different from conventional peer-assisted approaches for web content delivery, which mainly focus on alleviating the CDN servers' bandwidth load, we study how to use a browser-based peer-assisted scheme to resolve content delivery failures. Based on large-scale measurement studies on how users access and view webpages, we observe the challenges (e.g., peers stay on a webpage extremely short) that cannot be directly solved by conventional P2P strategies, and some important webpage viewing patterns. Due to these unique characteristics, WebRTC peers give us a novel way to deliver web content, coming with the problem of how to utilize the dynamic resources efficiently. We formulate the peer selection as an optimization problem, and design a heuristic algorithm based on the measurement insights to strategically select the most appropriate peer. Our prototype implementation on Tencent QZone and simulation experiments further demonstrate the effectiveness of our design: compared with non-peer-assisted strategy and random peer selection strategy, our design significantly improves the successful relay ratio of web contents under network failures, e.g., our design achieves a successful content download ratio of 60% even when users located in a particular region are not able to connect to the CDN servers.**

I. INTRODUCTION

Dominating today's Internet traffic, HTTP-based online services are suffering from many types of content delivery failures (e.g., server failures [1], network failures [2], routing failures [3]), significantly degrading the quality-of-service (QoS) provisioned to online users. This problem has been exacerbated when today's online services are designed to utilize aggregated contents and information, which are served by a geo-distributed Content Delivery Network (CDN), e.g., a web page consisting of contents that are served by multiple servers may fail to render at a client if it fails to download only parts of the contents from some servers.

To address this problem, we propose a joint CDN and peer-assisted web content delivery framework based on WebRTC [4], which has been widely supported by major browser vendors. Based on some observations from the measurment study including (1) high churning rate of online peers, (2) low probability of "re-join", (3) fragmentation of web content, we propose a novel peer selection strategy. By utilizing the

978-1-4673-7113-1/15 $31.00 ©2015 IEEE



(a) Conventional web content delivery.

(b) CDN+WebRTC web content delivery.

Fig. 1. Comparison between conventional content delivery and CDN+WebRTC content delivery.

WebRTC-powered browsers selected by our peer selection strategy to form "recovery" delivery paths, these widely deployed browsers can assist in content delivery when a user fails to receive requested contents from the original CDN servers, as illustrated in Fig. 1. Different from other WebRTC-based works [5], which mainly focus on alleviating the CDN servers' bandwidth load, we study how to use a browser-based peer-assisted scheme to resolve content delivery failures.

II. PATH-AWARE PEER SELECTION STRATEGY

Based on the measurement observations, we design a heuristic and distributed algorithm. We next present the factors learned from the measurement study and the detailed work-flows of our peer selection startegy, respectively.

1) Factors for Peer Selection: According to our measurement studies, content delivery failures show randomness with respect to the locations and time of requesting users. In our peer selection design, to improve the diversity of potential content delivery paths, we consider the following factors when choosing relay peers for a requesting user: (1) Diversity of peers' locations: In our design, we try to select relay peers who are at different locations to provide diverse candidate content delivery paths. (2) Diversity of peers' network (ISP): Similarly, we choose relay peers with diverse Internet Service Providers (ISPs) so that relay peers may have improved connections with the original content server under certain network failures. (3) Diversity of peers workload and *time-to-stay*: We also consider the relay peers' bandwidth consumption used by themselves and serving other requesting peers; meanwhile, we selectively assign a requesting peer with relay peers that have different time-to-stay, such that the relay peers are more likely to be online during the content delivery phase.

2) The Workflow of Peer Selection Algorithm: Based on the factors discussed above, when there is a relay request from peer, our peer selection algorithm works as follows. **Relay Peer List Generation:** (1) The requesting peer chooses relay peer among a candidate set, which contains potential relay peers, obtained from the coordinate server. The candidate set is generated based on the characteristics, discussed in the previous section, of each peer. (2) Ranking the list based on the workload and the *time-to-stay* of the candidate relay peers. **Failed Content Relaying:** The requesting user then actively tries these candidate peers to download the failed web content, i.e., the requesting peer first asks the *primary* peer (the first peer in the candidate relay list) to fetch the failed content, then it tries other candidates in the candidate list until the content is finally fetched.

III. Performance Evaluation

A. Prototype Experiments

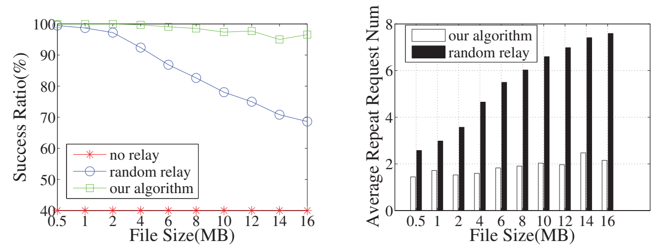
Using the algorithm in Section II, we implemented our design on Tencent QZone, and collected traces on how the algorithm works. The traces have the following information: (1) the requesting user identifier, (2) the relay peer identifier, (3) the timestamp that relay peer begins to relay, (4) the URL of the relayed content.

Based on these traces, we are able to detect and "recover" the failures of content downloads. In our traces, around 2% of content downloads encounter delivery failures, and our implementation is able to detect these failures and help the users download the contents.

B. Simulation Experiments

1) Experiment Setup: To verify the effectiveness of our design under extreme scenarios, we also carry out simulation experiments. In our experiments, we simulate $5,000$ peers that are distributed in five cities in China and randomly served by 3 ISPs. The peers are configured heterogeneously in terms of uplink and downlink capacities, following the statistics in [6]. We analyze the content size ranging from 500 KB to $16,000$ KB. **User behaviors:** We simulate the arrival patterns as a Poisson process with $\lambda = 30$, and the session duration as a Pareto distribution with $\alpha = 0.88$ and $b = 0.45$ according to our measurement study. **Content delivery failures:** In our experiments, we assume that users download web contents from regional CDN servers, and an in-network failure is the case that the network of a user's region is encountering some problems, and a fraction of users in this region may not be able to connect to the server.

2) Performance under In-Network Failures: First, we study the successful relay ratio, which is defined as the ratio of web content requests that can be finally served by CDN servers or peers. As illustrated in Fig. 2(a), we observe that our design can outperform the random peer selection strategy by 25%. Next, we study the impact of the in-network failure ratio on



(a) Average successful relay ratio.

(b) Average repeated request number in successful relays.

Fig. 2. Performance comparison under in-network failures

the successful relay ratio of the *primary* peer. We observe when the in-network failure ratio reaches 100%, i.e., all users in a particular region cannot connect to the CDN server, our design can still achieve a final successful download ratio of 60%. We omit the figure for saving space. We further investigate the average repeated request number in the successful relay session, which is defined as the number of repeated requests of requesting users to the candidate relay peers until it successfully downloads a content. In Fig. 2(b), we observe that compared with the random peer selection, users try a much smaller number of times to successfully download a failed content, especially when the file size is large.

IV. Conclusion

In this paper, we propose a joint CDN and peer-assisted web content delivery framework to address the content delivery failure problem. Our contribution lies in that we are the first to study to use a browser-based peer-assisted scheme to resolve content delivery failures. Based on large-scale measurement studies on how users access and view webpages, we not only show the challenges in our design that cannot be directly solved by conventional P2P strategies (e.g., peers stay on a webpage extremely short), but also learn webpage viewing patterns and design principles. We formulate the peer selection problem as an optimization problem, and design a heuristic algorithm based on the measurement insights to solve it.

Acknowledgement This work is supported in part by the National Basic Research Program of China (973) under Grant No. 2011CB302206, the National Natural Science Foundation of China under Grant No. 61472204, 61272231 and 61402247, the research fund of Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology and Beijing Key Laboratory of Networked Multimedia.

References

[1] "Facebook outage," https://www.facebook.com/notes/facebook-engineering/more-details-on-todays-outage/431441338919.

[2] D. Turner, K. Levchenko, A. C. Snoeren, and S. Savage, "California fault lines: understanding the causes and impact of network failures," *SIGCOMM Comput. Commun. Rev.*, vol. 41, pp. 315–326, 2011.

[3] D. Watson, F. Jahanian, and C. Labovitz, "Experiences with monitoring ospf on a regional service provider network," in *ICDCS*, 2003.

[4] http://www.w3.org/2011/04/webrtc-charter.html, W3C Real-Time Communications Working Group.

[5] L. Zhang, F. Zhou, A. Mislove, and R. Sundaram, "Maygh: Building a cdn from client web browsers," in *EuroSys*. ACM, 2013.

[6] M. Zhang, Y. Xiong, Q. Zhang, L. Sun, and S. Yang, "Optimizing the throughput of data-driven peer-to-peer streaming," *IEEE TPDS*, vol. 20, 2009.