Consolidated Review of
Peer-assisted Content Distribution in Akamai NetSession

1. Strengths

Unique dataset from a mature peer-assisted CDN in global deployment. The paper went to great lengths to explain the design goals and implementation of NetSession. I like how each analysis/result is geared towards answering specific question about the benefit or risk of NetSession.

I liked this paper a lot. The architecture is presented well, and while in hindsight they seem obvious design choices, it is impressive to see the scale at which it operates. I especially liked the bits about maintaining integrity using the infrastructure and throttling the peers upload bandwidth. Falling back on infrastructure nodes when needed gives a nice cushion to implement many of the neat ideas.

The authors' results indicate that hybrid CDNs can be practically deployed and at least some users will peer. Paper is well written and easy to read.

The paper presents a reasonably thorough measurement study of a live commercial system. The analysis is interesting, thorough and some of it yields surprising/novel insights. For example, the paper shows that NetSession does not affect the traffic balance of ISPs and thus the risk that ISPs will suffer is unfounded. The paper is well written and the graphics are clear.

2. Weaknesses

The bulk of the paper is tutorial in nature with regards to NetSession.

I did not get the sense that users were provided enough incentives. It seems like users simply stick to the default, and if the default allows peer assisting uploads, then the user is largely unaware of it. Even if the users sign the EULA I doubt they read it. The authors talk briefly about incentives and talk about how the content may be provided at lower cost, but it doesn't look like anything is being done about it. I think a better incentive structure will ensure continued success of this architecture.

NetSession focuses on static content rather than video, likely limiting its broad applicability to content delivery. Parts of NetSession have been published before, for instance, reference [1] by the authors analyzes NetSession form the client viewpoint. The paper didn't seem to give many new or unexpected insights.

The ideas in the paper are not particularly novel from an academic standpoint; rather the focus is on quantitative analysis of a mature commercial system. There are parts of the system where there is room to do creative things, such as in designing peer selection strategies for uploads and downloads. But the paper does not delve much into such issues.

The actual measurement/analysis results are actually quite slim for a long paper.

The paper is not very precise about defining the metrics it looks at and there are a lot of "speculations" about the results without further validation. Also, some of the results are not precise. There is a lot of "We speculate", and "It is possible that" etc. kinds of sentences

3. Comments

Overall, the paper included a great tutorial of the design and implementation of NetSession (a mature peer-assisted CDN that has been in deployment for 5+ years), as well as the various potential benefits and risks of such system.

I still don't understand what kind of content is being distributed in a peer-assisted manner. Clearly, Akamai is not using this to distribute content from CNN for example. It will be nice if you can give an example of the kinds of content being distributed. I wasn't sure why heterogeneity is a risk? The discussion about whether NetSessions increased AS load was unclear. You make the assumption that if ASes send the same amount of data between each other, they are in the clear. Doesn't it depend on the kind of peering? Also, only considering the source and destination AS really makes the analysis less useful. This section was the weakest in the paper.

- NetSession focuses on static content rather than video. The authors point out that this is one of their differences from reference [35], however they also point out in the end of Section 3.4 that this is partly because NetSession has had trouble getting video adoption. It would be helpful to have a better discussion of NetSession versus reference [35].
- How does NetSession ensure privacy? For instance, if a user gets content from a peer, that peer now knows the content requested by the user.
- The paper is well written but the writeup in section 3.7 on peer selection can be improved to better explain the two-level locality-aware peer selection strategy for downloads. Also, this reviewer could not find a description of peer selection strategy for uploads.
- From Table 4, two providers have >90% nodes with content uploads enabled, while five providers (the other extreme) have < 2%. Are they any particular characteristics about the content/objects that explain why a provider chooses an initial "enable" or "disable" setting?
- Section 5.1: the authors say that "peer-to-peer downloads were enabled for only 1.7% of the files, but these downloads accounted for 57.4% of the downloaded bytes overall." and "the average peer efficiency for peer assisted downloads was 71.4%." Can the authors give a better explanation of why the 57.4% and 71.4% are measuring different things?
- Figure 6 has extremely large error bars (e.g., ranging from 40% to 85% peer efficiency). Is there a reason for this large variability?
- In Section 5.2, the authors find that 25 to 30 peers are needed for good performance when downloading a particular file. How do these results compare, for instance to say bit torrent?
- In 5.2, the paper speculates that the discrepancy in performance associated with highest link bandwidths can be due to the asymmetry of upstream/downstream bandwidths. Given you have the IP address of clients, is it possible to investigate further the type of access networks the clients are connecting from to validate this conjecture?
- In Section 6.1, the information that's missing from Figure 10 & 11 is whether these AS-AS pairs truly have peering relationships or do they have customer/provider relationships

(which can be inferred from AS hierarchy using BGP data). For the latter, a traffic balance does not always mean "settlement" free.

- In the discussion about malicious peers (6.2), the paper merely states that NetSession relies on trusted edge servers to detect such attacks. Have such attacks been observed?

Finally, if you plan to release the measurement data, please put a pointer in the paper.

4. Summary from PC Discussion

This paper was accepted without discussion.

5. Authors' Response

We are grateful to the anonymous reviewers for their feedback. Our responses are below:

- **Incentives.** The NetSession team made a conscious decision not to include incentives, such as BitTorrent's tit-for-tat, and to instead serve content to each peer at the best possible speed. Disabling uploads does not adversely affect the user's download performance. We emphasized this more in Section 3.4.
- **Types of content**. At the time of writing, software installers were a typical kind of content; data files and other types, such as music and video, made up a small portion. We clarified this in Section 4.4.
- **Heterogeneity.** The main risks are the churn and the complex failure modes. We already state this in Section 2.4, and we elaborate further in Section 6.2.
- **Types of peering.** Figure 11 is based on direct links in the CAIDA topology data set, which does not distinguish between peer-to-peer and customer/provider links. We added a clarification at the end of Section 6.1 to emphasize that our results should not be interpreted as saying that NetSession's traffic has no impact on ISPs.
- **Privacy.** We agree that this is a possibility, but this information is never displayed to users and is deleted from the peer once the download completes and the logs have been uploaded to the infrastructure for billing and monitoring. We added a paragraph about this in Section 3.9.
- **Peer-selection strategy.** We added some further details to Section 3.7 of the paper, and we included a description of the uploading strategy in Section 3.4.
- **P2P enable/disable.** Some content providers choose to use peer assist, while others use NetSession simply as a download manager. We added a sentence to Section 5.1 to explain this.
- **57.4% vs 71.4%.** Suppose NetSession distributed only 1,000 bytes. Then 574 bytes would belong to downloads for which peer assist was enabled, and of those 574 bytes, 410 bytes (71.4%) would be sent by the peers. The infrastructure would send the remaining 164 bytes, plus the 426 bytes for infrastructure-only downloads. We clarified this in Section 5.1.
- **Large error bars.** Peer efficiency depends on a variety of factors, including the size of the object and the network connections of the peers that are contributing to the download in question. Hence, we can only hope to see a trend. We clarified this in Section 5.2.
- **25-30 peers.** These numbers are consistent with earlier studies of BitTorrent performance; we added references to two relevant papers in Section 5.2.
- **Asymmetric links.** Our geolocation data does include some rough estimates of access link bandwidth, but they do not distinguish between upstream and downstream bandwidth, so we were unable to validate the conjecture.
- **Data sets.** We regret that, in order to protect the privacy and confidentiality of Akamai's customers, we will not be able to make the data set available.