

Deep Diving into BitTorrent Locality

Ruben Cuevas
Univ. Carlos III de Madrid
rcuevas@it.uc3m.es

Nikolaos Laoutaris
Telefonica Research
nikos@tid.es

Xiaoyuan Yang
Telefonica Research
yxiao@tid.es

Georgos Siganos
Telefonica Research
georgos@tid.es

Pablo Rodriguez
Telefonica Research
pablorr@tid.es

ABSTRACT

Localizing BitTorrent traffic within an ISP in order to avoid excessive and often times unnecessary transit costs has recently received a lot of attention. In this work we attempt to answer yet unanswered questions like “*what are the boundaries of win-win outcomes for both ISPs and users from locality?*”, “*what does the tradeoff between ISPs and users look like?*”, and “*are some ISPs more in need of locality biasing than others?*”.

Categories and Subject Descriptors: H.4.3 Information Systems Applications: Communications Applications.

General Terms: Measurement, Performance.

Keywords: BitTorrent, ISP-friendship, Locality, Measurements, Peer to Peer.

1. INTRODUCTION

Several recent works [2, 6] have proposed architectures and protocols for localizing BitTorrent traffic. These works have looked at the problem of *how* to implement locality, but have not gone deeply into characterizing the conditions under which it is worthwhile deploying these technologies. The latter depends on the answer to several yet unanswered questions, including: (i) *Is locality a win-win for both ISPs and users, or does there exist a tradeoff between the two?*; (ii) *What are the main parameters affecting such a tradeoff and how do they vary across different ISPs?* and (iii) *Are some ISPs more in need of locality-biasing than others?*.

To answer the above questions we have conducted a large scale measurement study of BitTorrent demand demographics spanning 100K torrents with more than 3.5M clients at 9K ASes. We have also developed simple bounds on the performance of locality as well as scalable, yet accurate methodologies for computing traffic matrices from the above huge input without sacrificing essential BitTorrent mechanisms like the unchoke algorithm and the operation of seeders. We have validated our answers from the above study using an instrumented BitTorrent client and several live torrents.

A detailed description of the ongoing work introduced in this paper can be found in our Technical Report [3].

2. SPEED AGNOSTIC BOUNDS

We say that ISP A is on *sparse mode* with respect to torrent T if there do not exist many nodes outside A that

Copyright is held by the author/owner(s).
CoNEXT Student Workshop'09, December 1, 2009, Rome, Italy.
ACM 978-1-60558-751-6/09/12.

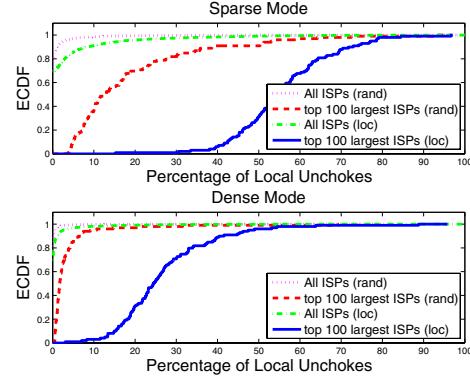


Figure 1: ECDF of Sparse (top) and Dense (bottom) metrics across all the ISPs in the dataset

participate in T and have similar speeds with nodes that are within A . Then due to *stratification* [5], nodes of A will be exchanging unchoke with each other but not with remote ones, even if the latter constitute the majority of their neighborhood. Similarly, we say that ISP A is on *dense mode* with respect to T if many remote nodes participating in T have similar speeds with the nodes of A .

The above definitions permit us to look at all the ISPs and torrents in our dataset and perform a simple probabilistic counting to compute the number of localized unchoke under sparse and dense modes for standard Random neighbor selection and a perfect oracle Locality policy. These extreme scenarios represent the bounds for Random (extreme sparse is the best case whereas extreme dense is the worst). Fig. 1 shows the obtained results: In sparse mode Random localizes 12.65% of unchoke in half of the top-100 ISPs. Locality on the other hand localizes 53.50% of unchoke. Thus Locality improves the median performance by a factor of approximately 4. In dense mode Random performs worse, localizing just 1.74% of unchoke in half of the top-100 ISPs; whereas Locality in dense mode localizes 24.40% of unchoke. The improvement factor of Locality in this case is around 14.

3. FACTORING IN THE SPEED OF ISPS

We define a new metric called *Inherent Localizability (IL)* that helps in understanding the impacts to a torrent under Random policy from real demand demographics (obtained from our own measurements) and ISP speed distributions ([1, 8, 7]). With this metric we get a more precise feel than

	LOIF	Locality	Strict
US1	19.86	34.07	95.82
US2	13.06	25.27	95.62
US3	12.96	23.45	95.14
EU1	9.18	34.77	95.66
EU2	8.60	36.88	94.82
EU3	24.84	42.71	96.05

Table 1: Transit Traffic Reduction in %.

with the previous bounds about the number of unchoke that can be localized in each case. We have computed the *IL* of two major ISPs in Europe (EU1) and US (US1). The *IL* of EU1 is generally higher than that of US1 for the same speed. This means that if the two ISPs had similar speed, then the demographic profile of EU1 would lead to a higher *IL* since this ISP already holds a big proportion of the content requested by its users. More importantly, we used *IL* to demonstrate that due to inhomogeneous demographics, speed distributions, and sizes of different ISPs, the amount of localized traffic changes non-monotonic with the speed of the local ISP. In other words, *becoming faster does not always help localizability*.

4. BITTORRENT TRAFFIC MATRICES

Our analysis up to now has been used for building up basic intuition on the parameters that affect the performance of Random and Locality. However it has a number of shortcomings (e.g. the analysis does not capture the behaviour of seeders and optimistic unchoke from leechers). In this section we use a more accurate model that addresses all these shortcomings and predicts the actual traffic matrix resulting from a set of torrents.

Computing Traffic Matrices:

We utilize fast numeric methods [4] that capture the unchoking behavior in steady-state. Notice that although experimentation with real clients would provide higher accuracy in predicting the QoS of individual clients, it wouldn't be able to scale to the number of torrents and clients needed for studying the impact of realistic torrent demographics at the AS level. Our scalable numeric methodology targets exactly that while preserving key BitTorrent properties like leecher unchoking (regular and optimistic) and seeding. We have validated the accuracy of our methods against real BitTorrent clients in controlled emulation environments and in the wild with live torrents (See [3]).

Locality biased Overlays:

We have defined a family of locality-biased overlays that captures the operation of existing overlay construction policies like the ones used in [2, 6]. Some notable members of interest in this paper are:

- *Local Only if Faster (LOIF)*: There is no constraint on the number of remote neighbors whereas switches of remote for local nodes occur only if the local ones are faster.
- *Standard Locality*: There is no constraint on the number of remote neighbors but local nodes are preferred independently of their speed to remotes.
- *Strict Locality*: All switches of remotes for locals are performed. Of the remaining remotes only one is retained and the rest are discarded.

Experiment Description:

We consider the following input to the experiments: (i)

	LOIF	Locality	Strict
US1	-2.24	1.69	12.93
US2	-1.81	1.03	16.58
US3	-2.96	0.03	21.82
EU1	0.62	6.22	24.48
EU2	0.88	6.08	13.33
EU3	-1.29	4.12	21.45

Table 2: Median QoS Degradation in %.

demands demographics from our large scale measurements and (ii) the speed distribution from [1](similar results have been obtained with other datasets [8, 7]). In our experiments we are interested in quantifying the effects of the described locality biased overlay construction on a “home” AS A. Thus, we compute the traffic matrices of all the torrents for AS A under the following policies: Random, LOIF, Locality and Strict. Out of the traffic matrices we define two metrics to be studied: (i) *transit traffic reduction compared to random* is of interest to the home AS; (ii) *user QoS reduction* (i.e. Download Rate Degradation) is of interest to the clients of home AS.

Summary of Results:

Table 1 and Table 2 present the transit traffic reduction and the user QoS reduction respectively for the 6 largest ISPs in terms of number of clients from our measurements (3 from Europe and 3 from US).

The main results obtained from our experiments are:

- The QoS preserving LOIF reduces transit traffic by around 20% in fast ISPs whereas in slow ones the transit saving is around 10%.
- Without firm constraints on the number of inter-AS overlay links, Locality achieves transit traffic reductions that top at around 35% in most of the ISPs that we have considered. The median QoS penalty on user download rate from Locality is typically smaller than 5%.
- The above bound on transit reduction is set by “unlocalizable” torrents, *i.e.*, torrents with one or very few nodes inside an ISP. Such torrents although amounting for around 80% of transit traffic under Locality, are requested by rather few users of an ISP ($\sim 10\%$). In a sense, the majority of users is subsidizing the few ones having a taste for unlocalizable torrents.
- By limiting the number of inter-AS overlay links huge reductions of transit ($\sim 95\%$) are possible. The median penalty is around 25%, whereas users on “unlocalizable” torrents can experience very high QoS penalties (97%).

5. REFERENCES

- [1] Ookla’s speedtest throughput measures.
- [2] David R. Choffnes et al. Taming the torrent: a practical approach to reducing cross-isp traffic in peer-to-peer systems. In *Proc. of ACM SIGCOMM ’08*.
- [3] Ruben Cuevas et al. Deep diving into bittorrent locality. Technical report, available from:
<http://arxiv.org/abs/0907.3874>.
- [4] Anh-Tuan Gai et al. Stratification in p2p networks: Application to bittorrent. In *Proc. of ICDCS’07*.
- [5] Arnaud Legout et al. Clustering and sharing incentives in bittorrent systems. In *Proc. of ACM SIGMETRICS ’07*.
- [6] Haiyong Xie et al. P4P: Provider portal for applications. In *Proc. of ACM SIGCOMM’08*.
- [7] Marcel Dischinger et al. Characterizing residential broadband networks. In *Proc. of ACM IMC ’07*.
- [8] Georgos Siganos et al. Apollo: Remotely monitoring the bittorrent world. Technical report, Telefonica Research, 2009.