Good Things Come to Those Who (Can) Wait

or how to handle Delay Tolerant traffic and make peace on the Internet

Nikolaos Laoutaris
Telefonica Research
nikos@tid.es

Pablo Rodriguez
Telefonica Research
pablorr@tid.es

ABSTRACT

Recent revelations that ISPs selectively manipulate P2P traffic have sparked much public discussion. Underlying this issue is the misalignment of interests between consumers on one hand who desire bulk transfers at flat rates, and ISPs on the other hand who are bound by budget and capacity constraints. Our thesis is that much of the tension can be alleviated by time-shifting traffic away from peak hours taking advantage of its *Delay Tolerant* (DT) nature. We propose two solutions for doing this. The first one offers incentives to end-users to shift their DT traffic and yet be compatible with flat-rate charging schemes. The second one posits augmenting the network with additional storage in the form of Internet Post Offices which can let ISPs perform store-and forward relaying of such DT traffic.

1. INTRODUCTION

The long term planning and deployment of infrastructure has always been a challenging task that requires predicting variables and future events that are unknown when the planning decisions are made – *"what will be the car usage in 5 or 10 years?"*, or, *"which areas in the vicinity of a large metropolis will develop more and thus require new roads and train connection to the city center'?'*. Similar questions are asked in the domain of networks for things like the future rate needs of residential and corporate connections, the dimensioning of backbones, and the peering agreements between ISPs.

Applications, Access, Backbone – Chasing the ever moving target: Much like in the previous examples taken from transportation, coming up with accurate predictions for the dimensioning of a network is a very hard task since there is too much uncertainty involved. For example, technological advances in access and backbone links often occur independently thus moving the bottlenecks anywhere between the end-user premises and the the network core [2]. At the application layer, the continuous introduction of new applications like P2P systems, user generated content websites, and multiplayer online games keeps changing the shape of network traffic matrices over increasingly shrinking time scales. Further, the difficulty of making accurate predictions is made worse by the fact that end-users are becoming increasingly involved in the introduction of new high bandwidth consuming applications and data. All the above points illustrate the difficulty of accurately predicting the future resource requirements of next generation networks. Therefore, bottlenecks are expected to keep appearing at one point of the network or the other and identifying them will continue being a chase of an ever moving target.

In such a volatile environment, it is important to have tools for relieving bottlenecks promptly and thus make time for network dimensioning to come up with more long term solutions. A prime objective of such tools would be to promote further the efficient usage of resources under the current triplet of applications, access, and backbone technology. Resource wastage – often referred to as "fat" in economics jargon – should be identified and removed promptly. But where can we find "fat" on the current Internet?

Delay tolerant applications and traffic: Consider the familiar example of a user who on receiving a suggestion, or after browsing a collection of media or applications, starts a large download that can take anywhere from a few to several hours. This is typically followed by additional time before the end-user really makes use of the information, e.g., watch the movie or install and start using the application. Such *Delay Tolerant* (DT) applications and their traffic allow much room for flexible scheduling and transmission, unlike interactive ones, like web browsing or video streaming, where requests and transmissions have to occur nearby in time.

DT applications therefore permit for a time-expansion of basic Internet scheduling and routing mechanisms. Internet routing has in the last few years gone through a *spatial-expansion* through technologies like overlay routing [3], anycast routing [4, 5], locality aware P2P network formation [1, 13, 6], etc. Scheduling, however, has not yet seen its own expansion, as it has been severely limited within the tight time scales imposed by congestion avoidance through TCP. The latter was designed under the overarching assumption that communication is interactive and intolerant to delay, which is not true for the aforemention class of DT applications.

As a consequence, both end-users and the network treat DT traffic like ordinary interactive traffic. Bulk downloads are initiated and accepted in the network during the hours of peak load despite the fact that the information they carry may be consumed several hours later. In the domain of transportation, such issues have been resolved through legislation. For example, in many places supply trucks are not allowed to make deliveries during commute hours, or access some highways during peak weekend traffic. On the Internet, however, there is no mechanism to prohibit DT applications from using limited resources during peak hours that interactive applications would value more. In that sense, DT traffic appears as "fat" in the pipes of ISPs.

Our contribution: In this paper we start by first identifying two basic causes behind the currently inefficient handling of DT traffic. The first one is the *lack of appropriate incentives for end-users* to self-select and schedule efficiently the transmission of DT traffic, e.g., postpone it until non-peak hours. This is a direct consequence of the prevailing flat-rate charging scheme that does not reward residential users that make efficient usage of network resources. Secondly, we point to a *lack of mechanisms on the part of the network* for identifying and handling DT traffic independently of how and when it is injected by the end-users. We propose two fixes with different pros and cons.

- Provide incentives under flat-rate charging: We argue that it is possible to keep flat-rate charging but still be able to incentivize the end-users to postpone their DT transfers until times of low utilization. The trick is to reward them for keeping their traffic low during peak hours, by providing them with bonus "higher-than-the-purchased" access rates during non-peak hours. For ISPs this makes sense since unutilized bandwidth costs nothing, whereas additional bandwidth during peak hours requires more investment in equipment.

- Allow the network to time-shift the DT traffic: We propose network attached storage in the form of *Internet Post Offices* (or IPOs) that will collect DT traffic in an opaque way from the end-users, and perform efficient transmission and scheduling based on the background load and the peering relationships between ISPs. We discuss two scenarios, one in which the local ISP operates the local IPO, and one in which IPOs are installed and operated by CDNs specializing in DT transfers. Such CDNs can become the catalyst for resolving tensions between ISPs and heavily consuming end-users.

Both solutions modify the flow of DT traffic, the first one at the source and the second inside the network. In the remainder of the article, we first discuss the impact of flat-rate charging on the way that end-users generate and transmit DT traffic, and then move on to elaborate on our proposals.

2. FLAT-RATE BROADBAND ACCESS

Despite the strong arguments [7] that economists have presented against flat-rate charging and in favor of more elaborate usage-based charging, flat-rate remains ubiquitous and has become a defacto standard for residential broadband access [10]. Undeniably, most of the appeal of flat-rate charging stems from its simplicity. It is easily communicable to the end-users who, in addition, feel safe by not having to worry about unpleasant surprises at the end of the month when the bill arrives, something not at all uncommon under usage-based charging schemes for other services like electricity and gas. For network operators, flat-rate charging obliterates the need to perform complex computations for calculating the charged amount of each user. On the negative side, flat-rate introduces the following problems.

- Unfairness: The common monthly amount that an ISP charges all users depends in the long run from the level of consumption of individuals and thus light users end up subsidizing the bandwidth of heavy users. When the difference between minimum and maximum consumption is not large, e.g., as in "all-you-can-eat" restaurants where the size of the human stomach puts very rigid bounds, then this is not much of a problem. In broadband access, however, as the rates increase, so does the maximum amount of unfairness due to cross-subsidy.

- Lack of incentives for efficient use of resources: Flat rate does not offer any incentives to end-users for making efficient use of network resources. Thus, even if a user knows that he won't be able to watch a movie until late at night or the weekend, there is no incentive for him not to start the download immediately. The reason is that postponing the download would place on the user the burden of having to remember to initiate it after the peak hours. Such wasteful usage habits combined with multimegabit Fiber-To-The-Home (FTTH) technologies can put an all too heavy strain on the infrastructure of an ISP. This partially explains why some ISPs have not yet released FTTH despite it being already a mature technology for the access.

There exists some partial solutions to these limitations. For example, usually one has the choice of multiple classes of flat-rate [7], each with different transmission rate and monthly cost. This however requires users to be able to predict accurately their bandwidth requirement and be willing to commit to it, as changing

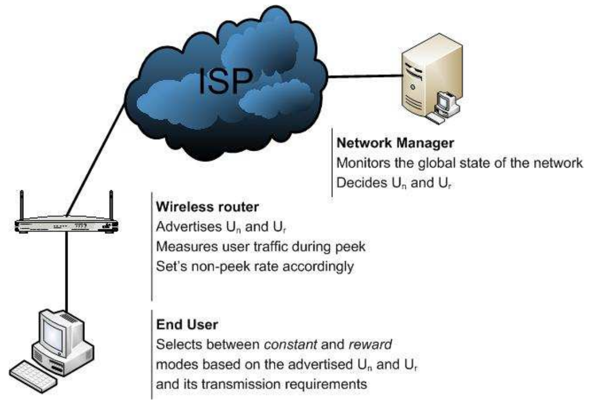

Figure 1: Architecture for implementing the *reward/constant* incentive scheme.

plans frequently based on usage habits is cumbersome. Similarly, some ISPs provide a capped download volume per month during peak hours and uncapped during off-peak hours. Although this allows for some flexibility (e.g., DT downloads can be put on crontab), it still ties the user to a particular daytime volume, and prohibits any kind of dynamic adjustment based on current usage habits. Unlike these two schemes, our proposal in the next section gives the end-user a very basic ability to modulate his available maximum rate according to his daily usage habits. Borrowing a term from the area of randomized algorithms, we will argue that network resource efficiency has much to gain from such an incentive scheme that embodies the *power of two choices*.

3. BUILDING INCENTIVES IN FLAT-RATE

In this section we show how to use the maximum allowed daily download volume as the means for building an incentive scheme into flat-rate charging.

3.1 Basic idea

A user pays a flat monthly amount for broadband access which entitles him to two different usage modes. In the first one (we will call it *constant*) the maximum allowed transmission rate has a constant value U throughout the duration of a day.[1] In the second one (we will call it *reward*) the maximum allowed transmission rate has value $U_n < U$ during the B "busy hours" of the network, and $U_r > U$ during the remaining $24-B$ hours of the day. The user can switch between the two modes on a day-by-day basis as will be explained next. *Reward* is designed to incentivize users to move all or part of their delay tolerant traffic away from the busy hours. The idea is pretty simple: by being "nice" to the network and keeping your rate below U_n (hence the subscript $_n$), you get "rewarded" with a higher rate U_r during the non-busy hours (hence the subscript $_r$). The values U, U_n, U_r, B must satisfy $U_n \cdot B + U_r \cdot (24-B) >> U \cdot 24$, i.e., permit a much higher overall daily transferred volume under *reward* than under *constant* with 100% utilization. P2P users with some ability for "Delayed Gratification" [11] would naturally respond to such a scheme.

[1]Henceforth, whenever we refer to the capacity of a link we mean the maximum of either direction.

The aforementioned example involving only 2 values (U_n, U_r) other than the standard one U, is the simplest possible reward scheme and as such it has the advantage of being the most easily explainable to the end-users. The idea, however, can certainly be generalized by making the non-standard rates a function of time, i.e., have $U_n(t)$ and $U_r(t)$ instead of constant values. In this case, the necessary condition for incentivizing the users to move their delay tolerant traffic away from the busy hours becomes: $\int_0^B U_n(t)dt + \int_B^{24-B} U_r(t)dt >> U \cdot 24$.

3.2 Architecture

The previous scheme can be fixed with respect to the values U, U_n, U_r, B, which would be decided once upon the establishment of a contract between a user and the ISP. It can be implemented very simply with the integration of minimal functionality on the user (PC) and ISP side (wireless router/gateway). For example, a simple button can be integrated to the user interface, allowing the end-user to select between *constant* and *reward*. Selecting the *reward* choice would set a self imposed cap of U_n during the busy hours through the OS and thus help the end-user meet the condition for receiving the *reward* rate during the non-busy hours. On the network side, all that is needed is to measure the transmission rate during the busy hours, and if it stays below U_n, then reward the user by increasing its allowed rate to U_r for the rest of the day. This is much simpler than trying to identify and shape DT traffic using elaborate deep packet inspection equipment. It leads to a win-win situation in which users are able to download more content, whereas ISPs do not need to over-dimension.

Another possibility is to keep only U fixed (going into to the contract) and communicate U_n, U_r, B dynamically to the end-user, letting him select accordingly. Fig. 1 shows the envisioned architecture. Such a scheme gives the ISP greater flexibility than the static one. For example, upon observing high utilization at some part of the network, the ISP can advertise lucrative "offers" for high U_r in an attempt to convince as many nearby users as possible to settle for a lower U_n.

One might argue that a similar scheme can be implemented only at the application layer, e.g., within downloaders and P2P clients, thereby obliterating the need for any kind of accounting on the network side. The problem of such an approach is that it cannot be enforced, as there will always be users that will hack the application and try to get U_r during the entire day.

4. ADDING STORAGE TO THE ISP: INTERNET'S POST-OFFICE

The previous incentive-based scheme requires minimal change in the infrastructure and the protocols used by an ISP. It rationalizes the use of network resources by incentivizing the end-users to time-shift their DT high rate transfers until appropriate times for the network. The price paid for not having to change the network, is that it requires end-users to pay some attention and e.g., decide whether they want to do P2P immediately or delay it to get higher daily volume. In the first case they would select the *reward* scheme through their user interface, otherwise they would continue with *constant*. In this section we look at ways to hide time-shifts from the end-users.

4.1 A storage enabled ISP architecture

In Fig. 2 we show a high level architecture for a storage enabled network involving the following two new elements. *Internet Post Offices* (or IPOs) which are just storage repositories at the access ISP, i.e., near the end-users. Since they are co-located, the IPOs can communicate with the end-users as fast as the access technology of the latter allows. There exists no other bottleneck or need for further investment to support such high rate transfers between the two.

Additionally, there exist *Transit Storage Nodes* (or TSNs) located at some PoPs at the backbone of the ISP, preferably near to peering-points with other networks. Of course, between end-user and TSNs, or between IPOs and TSNs, there can be all sorts of possible bottlenecks arising either due to congestion [2], or due to traffic engineering policies [12]. The key idea here, is to *use the IPOs and TSNs to time-shift bulk DT transfers, and thus avoid congestion and ISP-throttling, while making the time-shift transparent to the end-users.* The idea makes use of the fact that the price of storage is declining much faster than the price of bandwidth [8], especially at the access network. This approach is significantly different from previous attempts to tap on unutilized bandwidth (e.g., QBone's *Scavenger Service*[2]) that require changing the routers and cannot perform in-network time-shifting as they lack network attached storage.

4.2 A *fire-and-forget* approach to DT transfers

Imagine a user who wants to share with his friends a large collection of high resolution photographs and videos from his latest trip or vacations. Large numbers of such users having FTTH high rate access pose a formidable challenge to existing networks that are not dimensioned for such access rates and content sizes.

[2]http://qos.internet2.edu/wg/wg-documents/qbss-definition.txt

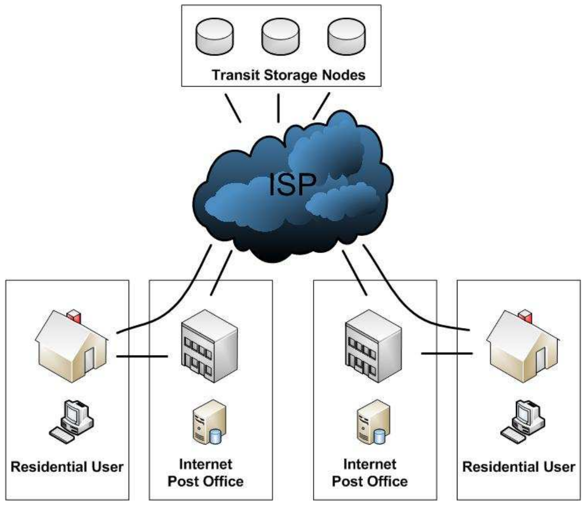

Figure 2: High level architecture of a storage enabled network.

Without substantial investment in upgrades of the backbone, the uplinks of DSLAMs, and the peering points to other networks, an easy solution for ISPs is to roll out FTTH and *police it heavily* when DT transfers like the above get into the way of servicing interactive, non-DT traffic. Of course, this would immediately trigger complaints from end-users expecting full FTTH rates at all times. Are there any other possibilities?

The aforementioned storage enabled architecture based on IPOs and TSNs suggests one. An end-user can push his collection of voluminous DT media to a local IPO at full FTTH rate. Since IPOs may connect directly to the DSLAMs, this does not put any strain on the rest of the network. Then the entry IPO can coordinate with other IPOs and TSNs to see that the collection reaches the intended recipients. This resembles snail (electronic) mail, where the end-user just hands in his mail to the local post office (SMTP server), at which point his direct involvement in the transfer comes to an end. A new breed of P2P applications can also be developed to make use of IPOs and TSNs. There are advantages from this for both the end-user and the network.

The end user: Benefits by pushing the data out of his computer at full FTTH rate. The end-to-end delivery has not been completed yet, but since the data are DT, what matters for the sender is how soon they will clear out from his computer and access line. After that, the user gets back his full CPU and uplink capacity for interactive tasks that would otherwise suffer from resource contention with slow and therefore long lived transfers. If the computer is a portable one, the user is free to disconnect and move. Last but not least, the user can shut the computer down much sooner, thus saving energy.

The ISP: Benefits by taking full control of the bulk transfer from the entry IPO and onwards, i.e., where most problems currently exist. The ISP can use IPOs and TSNs to schedule transfers at times of low background load. If the receiver is on an access network attached to the same ISP, then it can use the receiver's local IPO to bring the data down at a time of low utilization for the access network. If the flow has to cross to a different transit ISP, then this can be done when the corresponding peering point is least loaded.

5. CDNS AND DT TRAFFIC

The discussion up to now has been limited to ISPs and end-users. Next, we examine the potential gains for CDNs from handling DT traffic. We look at two scenarios based on the source of the DT traffic.

5.1 A CDN for Delay Tolerant Bulk data

Consider a CDN for servicing terabyte-sized *Delay Tolerant Bulk* (DTB) data, including scientific datasets, digitally rendered scenes from movie production studios, massive database backups, etc. Such a CDN installs storage nodes at access and transit ISPs from which it buys bandwidth according to a standard 95-percentile charging scheme [9]. Store-and-Forward scheduling is used to transfer DTB data between IPOs with the help of intermediate TSNs.[3] Our initial results based on real traffic traces from more than 200+ interconnection points of a large transit ISP show that SnF policies can reduce dramatically the transit costs incurred by End-to-End (E2E) policies that don't employ network storage. For example, with SnF we can transfer 100 Tbits of data from Latin America to Europe in 48 hours at zero transit cost, whereas an E2E stream of average rate of around 0.5 Gbps increases the monthly transit cost by tens of thousands of dollars under current bandwidth prices. The advantage of SnF lies on the fact that it can solve the problem of *non-coinciding load valleys* between the uplink of the sender IPO, and the downlink of a receiver IPO on a different time-zone. We explain the proposal through an example.

The top row of Figure 3 illustrate the 5-minute aggregate load on the uplink of an ISP in Latin America (LAT) hosting a sender IPO. The second and third rows depict the load on the downlinks of two ISPs in Europe (EU) and China (CH) hosting receiver IPOs. We have annotated with $uvalley(LAT)$ the time at which the uplink of LAT is least loaded and similarly for the downlinks of EU and CH. One can easily observe that due to time-zone differences, these valleys do not coincide. In the case of LAT and CH, the load valley

[3]For this example we have assumed that there are no bottlenecks inside the transit provider and thus it suffices to consider a single TSN.

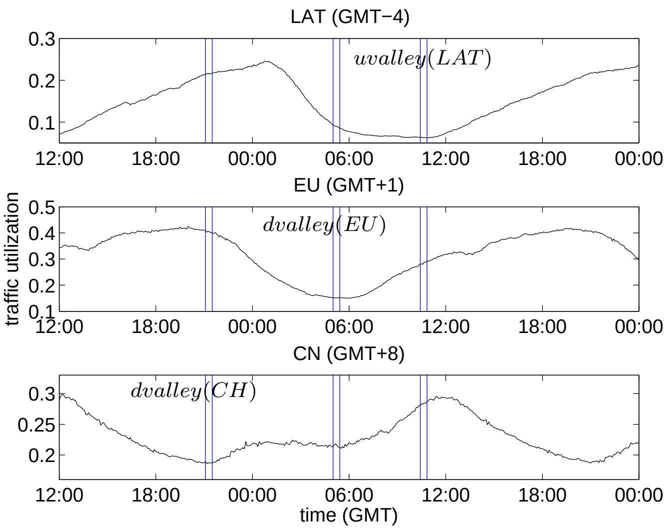

Figure 3: Time series plot of the uplink load of a sender in LAT and receivers in EU and CH. The uplink valley of LAT finds EU with substantial load and CH with peak load.

of the sender actually coincides with the peak of the downlink of the receiver. In this setting, E2E transfers will have to overlap with either a highly loaded uplink, or a highly loaded downlink and, thus, create either additional monetary costs by pushing the 95-percentile load based on which ISPs pay for transit, and/or obstruct the QoS of other interactive traffic with which the DTB traffic gets multiplexed. An SnF transfer through a TSN can do much better in this setting. It uses the uplink load valley to push data from LAT to a TSN on the transit ISP. The DTB data remain buffered there until the beginning of the load valley of the downlink of the receiver, at which point they are pushed to their final destination (EU or CH).

Examining all the pairs from the 200+ peering points of our transit provider we found that more than 50% of the busiest pairs had valleys that were apart for at least two hours, and thus cases like the aforementioned example were not at all uncommon. Non-coinciding valleys appear frequently, even within the same or nearby time-zones. This happens because networks of different type can peak at different hours, e.g., a corporate network typically peaks during work-hours, whereas an ADSL access network typically peaks in the late evening.

5.2 A CDN for Delay Tolerant End-User data

Next we look at what CDNs can do for residential end-user DT traffic. The model is similar to Fig. 2 with the difference that storage nodes are not managed by the ISP, but by an independent CDN which, unlike the ISP, has global coverage with PoPs on multiple networks. Again IPO nodes are used for collecting end-user DT data at full FTTH rate, whereas other intermediate IPOs and TSNs help complete the delivery. The CDN can sell this service to content creators and give it for free to end-users. In addition to its obvious benefits for content creators and end-users, the operation of such a CDN adds value to the ISPs. The reason is that since it receives end-user DT traffic, the CDN can transmit it in an ISP-friendly manner unlike most end-user applications. For example the CDN can:

Prefer peering to transit links: Having post offices at multiple ISPs, the CDN can try to create an end-to-end path between a sender and a receiver that involves mostly peering links between neighboring ISP, over which traffic is exchanged without monetary transit costs. Transit links can be used only in cases that alternative paths through peering links do not exist or are severely congested.[4]

Avoiding the hours of peak load: The CDN can take advantage of the DT nature of the traffic to avoid times of high utilization. In the case of transit links this protects against increases of the 95-percentile of send traffic, and corresponding increases of monthly charging bills. In the case of peering links, it preserves the QoS of the background traffic and avoids the need to upgrade the link and incurring additional equipment and maintenance costs.

We believe that establishing and demonstrating the above practices would permit a CDN operator to achieve a symbiotic relationship with the ISP. The ISP would benefit by having the heat of end-user DT traffic taken away from it thanks to the CDN. The CDN would benefit by obtaining cheap flat-rate or even free access to ISP bandwidth under the conditions of ISP-friendliness discussed above. Notice that unlike Sect. 5.1 in which the CDN was introducing new *exogenous* DTB traffic to the ISP, and thus had to pay according to 95-percentile charging for it, now the CDN is just servicing *endogenous* end-user DT traffic, including high definition video from P2P, that already flows in the ISP.

6. CONCLUSIONS

In this article we claim that many of the tensions that currently exist on the Internet are due to wasteful usage of resources during the hours of peak load. A first step towards reducing such wastage is to shift what we define as Delay Tolerant traffic to non-peak load hours. We have proposed two solutions for this, one by offering *flat-rate compatible incentives to the end users*, and a second one based on the addition of *network attached storage*. The first solution has the advantage of requiring minimal change to the existing network, but requires a small involvement from the end-users. The second solution is completely transparent to the end-users but requires the addition of network attached storage. The latter proposal becomes economically efficient since the price of storage has been declining much faster than the price of network equipment, especially at the access and regional network.

[4]Notice that although no immediate transit cost is paid for crossing peering links, there still exist implicit, albeit real costs. For example, if the peak utilization becomes too high due to the additional delay tolerant traffic then the ISPs will have to upgrade the speed of their peering and thus incur capital and maintenance costs in order to preserve the QoS offered to their clients. We consider this next.

7. ACKNOWLEDGEMENTS

This work is supported in part by the NANODATACENTERS program (FP7-ICT-223850) of the EU. The authors would like to acknowledge the help of P. Chhabra, V. Erramilli, G. Smaragdakis, R. Sundaram as well as the anonymous reviewers in improving the paper.

8. REFERENCES

[1] Vinay Aggarwal, Anja Feldmann, and Christian Scheideler. Can ISPs and P2P users cooperate for improved performance? *ACM SIGCOMM Comput. Commun. Rev.*, 37(3):29–40, 2007.
[2] Aditya Akella, Srinivasan Seshan, and Anees Shaikh. An empirical evaluation of wide-area internet bottlenecks. In *Proc. of ACM SIGMETRICS '03*, pages 316–317, 2003.
[3] David Andersen, Hari Balakrishnan, Frans Kaashoek, and Robert Morris. Resilient overlay networks. In *Proc. of ACM SOSP'01*, Banff, Canada, Oct 2001.
[4] Hitesh Ballani and Paul Francis. Towards a global IP Anycast service. In *Proc. of ACM SIGCOMM'05*.
[5] Hitesh Ballani, Paul Francis, and Sylvia Ratnasamy. A Measurement-based Deployment Proposal for IP Anycast. In *Proc. of ACM IMC'06*, October 2006.
[6] David R. Choffnes and Fabián E. Bustamante. Taming the Torrent: A practical approach to reducing cross-ISP traffic in P2P systems. In *ACM SIGCOMM'08*.
[7] Costas Courcoubetis and Richard Weber. *Pricing Communication Networks: Economics, Technology and Modelling*. Wiley, 2003.
[8] Bradford DeLong. Shifting into overdrive, what happens when mass storage leaves microchips in the dust. *WIRED Magazine*, (11), 2005.
[9] David K. Goldenberg, Lili Qiuy, Haiyong Xie, Yang Richard Yang, and Yin Zhang. Optimizing cost and performance for multihoming. In *Proc. of ACM SIGCOMM '04*, pages 79–92, 2004.
[10] Andrew M. Odlyzko. Internet pricing and the history of communications. *Computer Networks*, 36:493–517, 2001.
[11] Y. Shoda, W. Mischel, and P. K. Peake. Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology*, 26(6):978–986, 1990.
[12] Slashdot. Comcast Hinders BitTorrent Traffic, Aug 2007.
[13] Haiyong Xie, Yang Richard Yang, Arvind Krishnamurthy, Yanbin Liu, and Avi Silberschatz. P4P: Provider portal for applications. In *Proc. of ACM SIGCOMM'08*.