

# An Internet Coordinate System to Enable Collaboration between ISPs and P2P Systems

Vinay Aggarwal, Anja Feldmann, Roger P. Karrer

Deutsche Telekom Laboratories / TU Berlin  
{vinay.aggarwal, anja.feldmann, roger.karrer}@telekom.de  
Ernst-Reuter-Platz 7, 10587 Berlin, Germany.

## Abstract

P2P systems, which contribute a significant portion of today's Internet traffic, build their overlay topology largely agnostic of the Internet underlay. This leads to traffic management challenges for ISPs on the one hand, and potentially inefficient neighbourhood selection for P2P nodes on the other hand. To alleviate this problem, we propose a novel global Internet coordinate system, which can be used by P2P and other applications, to get estimates about path properties to potential neighbours/servers, both within and outside their ISPs. These estimates can be used, e.g., by P2P users to pick appropriate neighbours, so that both ISPs and P2P systems benefit. The coordinate system is built through ISP-P2P collaboration on the one hand, and collaboration between multiple ISPs on the other hand. In this paper, we also discuss which metrics can be provided by our coordinate system, and how it differs from existing coordinate systems.

## 1. Introduction

Over the last decade, peer-to-peer (P2P) systems have become one of the dominant applications in the Internet. Their traffic already contributes more than 50% of today's Internet traffic [1]. The P2P concept is used by a multitude of different applications: file sharing as in Bittorrent, eDonkey, Kazaa, and Gnutella; real-time multimedia streaming such as MySpace, Yahoo, and YouTube; or phone systems using VoIP such as Skype, MSN, and GoogleTalk. These applications profit from the almost unlimited scalability properties of P2P systems. Indeed, P2P applications are even one of the main reasons for customers to upgrade to broadband access and are therefore revenue boosters for Internet Service Providers (ISPs) [2].

This paper investigates the relationship between P2P systems and ISPs. We show that this relationship lacks coordination and is therefore tense. Hence, we advocate enabling collaboration between ISPs and user applications as well as between multiple ISPs, and propose to use a global coordinate system to overcome such problems.

## 1.1 Need for collaboration between ISPs and P2P systems

The main characteristic of P2P systems is that they build an overlay on top of the Internet infrastructure. This overlay is used to connect peers in an overlay topology and to forward data along overlay paths. We argue that unfortunately, current P2P systems are inefficiently built and operated, to the disadvantage of both ISPs and P2P systems. The reason is that P2P systems build and operate their topology *independently* of the underlying network topology. Neighbours and paths are often chosen at random in unstructured P2P systems, while they may be based on some metrics such as round-trip times (RTT) in structured P2P systems. Thus, it is frequent that neighbours in a P2P network are located in different ASes.

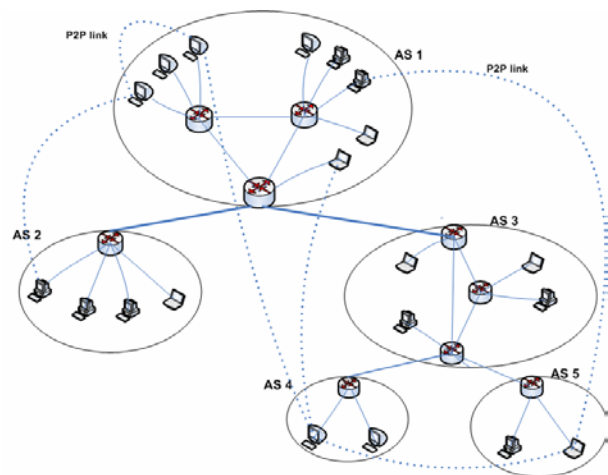


Figure 1: Relationship between P2P users and the Internet AS topology

Figure 1 shows a very simple Internet topology, consisting of 5 Autonomous Systems (AS). Each AS consists of border routers, internal routers and end system hosts. The dotted lines correspond to overlay links between two nodes in the P2P system, and two peers connected via such a link are considered P2P neighbours. As depicted in Figure 1, two nodes that are neighbours in a P2P system may be physically located

in two different ASes (e.g., AS 1 and 5). These ASes may be as far away as Europe and Australia. The actual path between P2P neighbours in this case goes through AS 3, crossing multiple AS boundaries and access links, which can account for significant performance penalties in terms of available bandwidth and latency. Hence, the notion of neighbourhood can significantly differ if seen from a P2P node rather than an ISP's viewpoint.

The drawback for the P2P system is obvious: distances in terms of RTT as well as connectivity in terms of bandwidth and reliability can be dismal. In particular, unstructured overlays show poor performance, and in structured overlays, the measured RTT does not always correspond to peers being well connected in terms of bandwidth [19].

For ISPs, the independence of P2P systems has a two-fold drawback. First, P2P traffic crosses multiple links and multiple ASes. This crossing incurs significant costs for ISPs [16] – which is crucial in times where flat rates for Internet connectivity have destroyed the relationship between user traffic and ISP revenues. Second, ISPs have difficulties engineering their traffic. The overlay network of a P2P system duplicates the routing control functionality at the application layer that is already available within the routing layer. Thus, with P2P systems, the Internet has two control loops. Because they are independent and not coordinated, they interact adversely. For example, assume that an ISP shifts load from one link to another to free up resources. The P2P system may notice this shift and interpret it as an increased congestion on one link and a light traffic load on the other. It adjusts its overlay routing strategy accordingly – and thereby inverts the efforts of the ISP!

The lack of control over P2P behaviour makes it difficult, if not impossible, for ISPs to control and engineer their traffic [2, 16]. Therefore, P2P as well as all other user traffic, including Web and VoIP, may suffer quality degradations which can be especially drastic at peak loads. This is devastating for ISPs who wish to offer good service to all customers. Therefore, we argue that ISPs and P2P systems should collaborate, for their mutual benefit.

## 1.2 A Coordinate System

The lack of control and collaboration between P2P applications and ISPs can be addressed in two ways. The first solution is that a P2P system can try to infer the location and the distance to other peers on its own. A plethora of tools are available, see e.g., [13, 14], that allow an end host to estimate the topology, delay, bottleneck or the available bandwidth at different paths. However, active measurements impose a significant overhead and, when used by many peers, are far from being a scalable solution [15]. Moreover, active

measurements do not solve the problem of conflicting interactions at the routing and application layers.

The alternative is that the P2P network uses the services of a system that provides network distance information about other peers to a querying application node, such as a coordinate system [3]. A coordinate system maps the IP address of a peer into an n-dimensional coordinate space. The coordinate distance between two nodes in that space reflects the network distance between them in the Internet, which is defined as the RTT propagation and transmission latency.

We believe that a P2P application can use the coordinate system in two ways. First, it may use the system to get an estimate of the network path properties between any two nodes in the system. Second, a peer may submit its own address and a list of potential neighbour peers, and ask the coordinate system to sort the list in increasing order of their distance to itself. Using these functions, an overlay topology of a P2P system can be built that reflects the real distances in the physical topology. In particular, nodes should only be neighbours in a P2P system if their distance in the coordinate system is small.

Coordinate systems have been proposed previously [3–7], however, we argue that the way such coordinate systems are built are again far from efficient and suitable. Existing coordinate systems measure the RTT and map the distances into a low dimensional Euclidean system like the Cartesian coordinates [3], or a non-Euclidean one, e.g., hyperbolic, spherical, or toroidal [7]. Unfortunately, the actively measured RTTs are far from accurate and may change quickly over time [19]. Moreover, up to now, they cannot offer available bandwidth or capacity estimates.

## 1.3 Our Proposal

Therefore, we propose an alternative way for building a coordinate system: namely by collaboration among ISPs. ISPs have detailed information about the connectivity of peers that are located within their domain: their bandwidth, their usage patterns, etc. Moreover, ISPs also decide and implement their routing policy, and are thus aware of the routing paths within their network and to other ISPs. By using this already available ISP information and exchanging summaries of it among ISPs, a coordinate system can therefore be built that does not require active measurements. Moreover, we argue that this coordinate system is more accurate, is capable of addressing additional metrics and can provide the information quicker to a querying node than current coordinate systems.

In Section 2, we elaborate on this concept, and explain how we build a coordinate system for a single ISP,

utilizing the oracle server introduced in [1]. We then describe the information exchange between multiple ISPs to build a global coordinate system in Section 3. This is followed by a discussion on related work in Section 4, and a comparison of our proposed system to existing coordinate systems. Finally, we conclude in Section 5.

## 2. A coordinate system for a single ISP based on ISP-P2P collaboration

This section describes our approach for building a coordinate system for a single ISP, based on collaboration between an ISP and P2P or other user applications running within the ISP network.

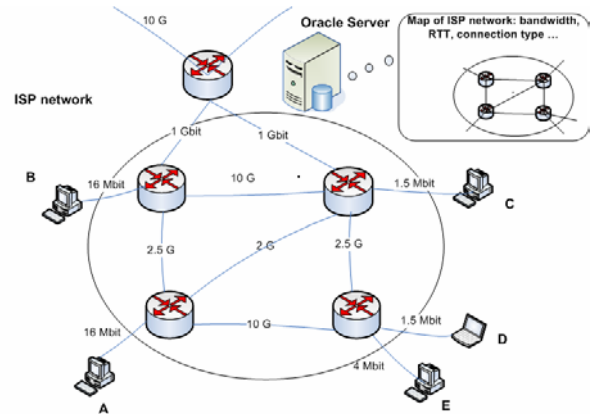
The basic task of a coordinate system is: given two IP addresses return an estimate of the network distance (usually defined in terms of RTT) between them. Our main insight is that ISPs either have or gather the most relevant as well as accurate information about the connectivity of hosts that are located in the ISP's domain, where the term connectivity includes information such as physical bandwidth to the last hop (modem, DSL, VDSL, etc.), latency statistics, geographical location and customer service class including different quality classifications, such as gold, silver, or normal customer. Moreover, each ISP decides the routing policy for transmitting traffic within its network, using intra-domain routing protocols like OSPF, IS-IS, and RIP. In other words, an ISP is already in possession of the information that other coordinate systems have to infer including link capacity, service classes, available bandwidth, estimated delay/RTT, etc. Hence, given two IP addresses within its network, an ISP can determine or estimate a summary of the basic path characteristics of the network path between them.

### 2.1 The oracle concept: a coordinate system for a single ISP

We, in [1], introduced and evaluated the idea that each ISP hosts an oracle server. The oracle has access to an up-to-date map of the ISP network - it knows the bandwidths of the links within the network, the connectivity characteristics of the users, and the estimated RTT, see Figure 2. The oracle offers an information service to applications within its network. For example when a P2P user wishes to select a neighbour or decide from which neighbour to download a file from, it can supply the oracle with a candidate list of peers. The oracle then ranks this list according to their proximity to the querying node, which can then be used by the P2P node to select a near neighbour. The ISP can use this mechanism to steer P2P traffic, for example, to express preference for P2P neighbours that

are within its network. The oracle can rank the candidate list based on a number of factors:

- nodes within the AS network are preferred
- intra-domain routing information, like OSPF metrics, geographical location or last-hop connection characteristics of customers, e.g., DSL, modem, etc. are used to fine-tune the list of nodes within the AS.



**Figure 2: Oracle server provides a coordinate system for a single ISP.**

Consider the example network shown in Figure 2. It shows the internal topology of an ISP, with various users A, B, C, D, and E having different connection bandwidths at the connection edge. When user A wishes to connect to another peer for bootstrapping to a P2P network, we assume that it finds B and E as possible candidates through a P2P bootstrapping mechanism, e.g., a Web Cache or previously stored list of active users. Now, A queries the oracle server for path properties of B and E. The oracle server knows that B has a last hop bandwidth of 16 Mbit, which is much larger than the 4 Mbit bandwidth of E. Hence, it recommends A to connect to E. The oracle can either rank B ahead of E, or can return a bandwidth classification of B as high, and E as medium. This enables A to connect to a user having a better bandwidth. Consider another instance, when D is already connected to both E and C in a P2P network. When D wishes to download a large multimedia file, it queries the oracle about its connected neighbours. The oracle can tell D that E not only possesses a better last-hop bandwidth, but is also closer geographically and topologically.

The oracle service can be realized as a single server, or as a set of replicated servers within each ISP, that can be queried using a UPD-based protocol or can run as a Web service. The exact implementation details can be decided by each ISP independently.

We can see that the proposed oracle service already provides an abstraction of a coordinate system. In the

terminology of current coordinate systems, each ISP – represented by its oracle server – is the pendant to a landmark. However, instead of measuring distances between different landmarks and between landmarks and peers as is the case in existing coordinate systems [3-7], each ISP’s oracle stores connectivity information to build a coordinate system. Compared to existing coordinate systems, it has a number of advantages. First, the knowledge of the oracle goes far beyond knowing the distance between peers in terms of RTT only. With its knowledge about link capacities, available bandwidth, geographical location, etc., it can also answer questions such as “which peer has the best bandwidth to me?” Even combinations of multiple metrics are possible.

A critical issue in coordinate systems is privacy. A user that has a broadband connection may not be that excited about getting requests from thousands of peers. Moreover, if the ISP is behind the coordinate system, the question must be addressed if the ISP is actually allowed to provide this information. Our answer to privacy concerns is threefold. First, the information revealed to a user by the ISP can anyway be estimated using tools such as [13, 14]. Second, the ISPs and the oracle are not requested to provide all the details of the connectivity. An oracle can keep the details of why a list of peers is ranked in a particular way confidential. Moreover, it can create a simplified classification of connectivity, such as good, medium and low, and just provide these classifications to users or other ISPs. Finally, an ISP may alter ranking criteria dynamically. For example, if 100 queries always include the same peer, the ISP may decide to announce a lower connection property for this peer to shed load.

## 2.2 ISP-P2P collaboration

The authors of [1] have shown the benefits and advantages of using the oracle for both P2P systems and ISPs. Recall, that P2P systems often choose their neighbours without any respect for network locality, that the P2P traffic often crosses ISP boundaries multiple times, even though the desired content is often available in the proximity of interested users. P2P nodes normally select a peer to connect to, or to download content from, by choosing from a list of available peers, based on some degree of performance measurement or agnostically.

The P2P and other application users benefit from the oracle service because they do not have to measure the path performance themselves, they can take advantage of the ISP’s knowledge and they are able to avoid low-latency paths or bottlenecks at inter-ISP transit/peering links, thus experiencing better performance in terms of faster response times and better available bandwidths. That P2P users benefit from consulting the oracle, both

while bootstrapping to the P2P network, as well as while selecting a peer to download content, has also been shown in [1].

The ISP gains by regaining control of the P2P traffic. For example it can now avoid that a large portion of the P2P traffic leaving its network, thus implying enormous cost advantages [16]. Also, the ability to route large amounts of P2P traffic along desirable links improves the ability of the ISP to offer better QoS to P2P as well as HTTP and other traffic, thus leading to better customer satisfaction.

While there are some proposals that aim to localize P2P traffic [17, 18], the oracle solution is simpler and more general. It is applicable to P2P nodes of all overlays, and also promotes collaboration between ISPs and P2P users, who are otherwise found to be at loggerheads with each other. Also, the significant overhead of reverse-engineering the path properties through extensive Internet measurements [15] is eliminated, a relief for both ISPs and application users.

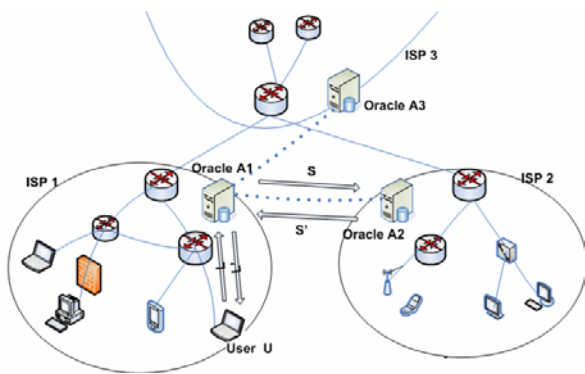
We now show how we leverage the oracle concept to build a global coordinate system, based on collaboration between multiple ISPs.

## 3. A global coordinate system through multiple ISP collaboration

Not only does each ISP know its own network, it also knows its routing policies to other neighbouring ISPs. Each ISP has information about which of its neighbouring ISPs are customer, peer or provider ASes. As it routes traffic to and receives traffic from other ASes, it also has BGP [20] path information to other ASes, and a fairly good estimate about the IP address ranges of their customers. Besides, an ISP is also aware of the capacities and other characteristics of inter-domain backbone links, at least in its neighbourhood. We propose to use this vast information available at each ISP to build a global coordinate system, by extending the oracle service outlined in the previous section. We envisage a system where the oracle servers from the different ISPs collaborate, see Figure 3.

When an oracle receives a list of candidate IP addresses from a user within its network, it can rank the nodes within its network using its network information, as described in Section 2. For nodes that do not belong to its own network, it first segregates them according to their parent ASes. Nodes that belong to ASes in its immediate neighbourhood can be further classified as belonging to customers, peers or provider ISPs. As each ISP has to pay for traffic going to upstream provider networks, it has an interest in preferring nodes from customer and peer links, depending on its individual

AS-level routing policy. Hence, for instance, customer and peer-ISP users can be higher ranked than provider ISP's users. If the queried IP addresses do not belong to ISPs in its immediate neighbourhood, an ISP can also use AS-hop distance, defined as the number of AS-hops on the chosen BGP route for the IP address, or BGP routing policy [20] (preferred AS paths, point-of-exit for traffic, etc.) for ranking nodes outside its network. For instance, nodes belonging to ASes with lesser AS-hop counts will be ranked higher as compared to nodes belonging to ASes farther away. The level of granularity for ranking the list of nodes can be decided by each ISP independently. To further fine-tune the list of nodes, the oracle contacts the oracle server from the neighbouring ISP, and sends it the list of users that it wants to rank.



**Figure 3: Communication between oracles of different ISPs to make a global coordinate system**

Consider the scenario in Figure 3. ISP1 is connected by a peering link to ISP2, and by an upstream link to ISP3. When oracle A1 gets a list  $L$  of candidate IPs from a user  $U$ , it sorts the list of IPs within ISP1 on its own, using a semi-static database containing information about its network. As ISP1 prefers to route traffic to ISP2 instead of ISP3, it estimates the subset  $S$  of the list of IPs which belong to ISP2, and sends it to the corresponding oracle server A2. The oracle A2, on receiving this list  $S$  of IPs which belongs to its network, can easily rank this list based on metrics like available link capacity, estimated delay, geographical location, etc. as described in Section 2. It then sends the ranked list  $S'$  back to A1, which A1 incorporates into its final ranked list  $L'$  to be returned to the user  $U$ . Depending on the level of fine-tuning desired, A1 can even contact multiple neighbouring oracles, e.g., A3.

Hence, each ISP, using a combination of ISP-P2P collaboration (for individual network), and ISP-ISP collaboration (for multiple networks) can provide a global coordinate system. Please note that each ISP is free to rank the list of IPs according to its own criteria and has full control over how much information it wishes to reveal. An oracle can simply rank the list of nodes, or can classify nodes into different classes.

## 4. Related Work

In recent years, research on Internet coordinate systems has received much attention. Most of the coordinate systems proposed so far including [3,4,5,6,7] rely on a set of landmarks or on peer-to-peer technology. The current set of coordinate systems mainly attempt to map hosts into synthetic coordinates in some coordinate space (Euclidean [3], spherical [7], hyperbolic [7] being some examples) such that the distance between two hosts' synthetic coordinates reflects or estimates the actual round-trip-time (RTT) between them in the real Internet. While this approach may serve well for some applications like Web servers or content distribution systems (CDNs) which do not experience high churn, its leads to performance degradation in the face of newer brands of file sharing and CDNs which are characterized by high user churn [12], pollution in content and malicious activity on the part of the users. Coordinate systems are vulnerable to malicious users who lie about their locations [9, 10]. They have also shown that coordinate systems are several orders of magnitude slower than direct ping measurements made by individual peers, often taking several tens of seconds or even minutes to converge. This is clearly unacceptable given the high churn in P2P systems, and their small online durations [12]. Lua et.al. [8] have shown that the accuracy metrics used by these coordinate systems are not accurate enough. More recently, Ledlie et.al. [11] have shown that while the performance of the coordinate systems reaches expected levels on Planetlab nodes and simulation environments, the performance degrades significantly when deployed in the real Internet. Recent studies [19] have also shown the limitations of using RTT as a metric for coordinate systems. The existing coordinate systems predict network distance as a sum of RTT propagation and transmission delay, which they assume to be a fairly stable characteristic between Internet hosts. However, RTT is dependant on network load, which is heavily influenced by factors like churn and bursts in user activity in P2P and CDN systems. As such systems are dominated by peers who have very short uptimes [12], assuming RTT to be a stable metric is not a sound assumption.

### 4.1 Discussion

Compared to the above systems, our proposed coordinate system does away entirely with the complex mathematical computation process of mapping a node's location in the Internet to a point in the mathematical coordinate space. As the node location, its connection information and the network routing policy is known to the oracle, the need for Internet measurements [15] and parameter estimations is heavily reduced, thus reducing network overhead and increasing scalability. Also, our system is not based solely on RTT. Rather, the network

distance between two nodes reflects not only the RTT propagation and transmission delay, but also factors like:

- path capacity and available bandwidth
- better paths which may or may not correspond to least RTT, but do offer better bandwidth and packet loss rates
- respect for AS relationships, BGP-based policy routing and other routing metrics like point-of-exit of AS, next-hop AS, multi-exit discriminator (MED), etc.

That the oracle system is resistant to churn in P2P applications has already been demonstrated in [1]. As the oracle does not need to ask nodes for their location, the susceptibility of the system to malicious nodes lying about their location is also removed, a major improvement over existing coordinate systems. Moreover ISPs can tailor their answers to regain control over their traffic.

## 5. Conclusion

We have proposed a novel global coordinate system, built on the concept of collaboration between user applications and ISPs on the one hand, and between different ISPs on the other hand. The proposed system provides accurate network distance information, using not only RTT, but also other important metrics like path capacity, available bandwidth, customer service class, AS relationships and routing policies, without the need for reverse-engineering the Internet by large scale measurements. The system is scalable, resistant to churn, and less susceptible to malicious nodes. The coordinate system can be used by all kinds of user applications, which need some estimation of network properties to choose appropriate neighbours. With the Internet transforming from a client-server model to a user-generated-content model, where different nodes generate, search, seek and download content at the same time, and where the content ranges from low negotiation traffic to heavy multimedia content, such a scalable and accurate global coordinate system can contribute significantly to both ISPs and Internet users.

## References

[1] V. Aggarwal, A. Feldmann and C. Scheideler. Can ISPs and P2P Users Cooperate for Improved Performance? In ACM SIGCOMM Computer Communication Review, 37(3), July 2007.

- [2] T. Mennecke. DSL Broadband Providers Perform Balancing Act .Online: <http://www.slyck.com/news.php?story=973>.
- [3] P. Francis, S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavitt and L. Zhang. IDMaps: A Global Internet Host Distance Estimation Service. In IEEE/ACM Transactions on Networking, Oct 2001.
- [4] T. Ng and H. Zhang. Predicting Internet Network Distance with Coordinates-Based Approaches. In IEEE INFOCOM, 2002.
- [5] F. Dabek, R. Cox, F. Kaashoek and R. Morris. Vivaldi: A Decentralized Network Coordinate System. In ACM SIGCOMM, 2004.
- [6] H. Lim, J. Hou and C. Choi. Constructing Internet Coordinate System Based on Delay Measurement. In IEEE/ACM Transactions on Networking, 13(3), 2005.
- [7] Y. Shavitt and T. Tankel. On the Curvature of the Internet and its Usage for Overlay Construction and Distance Estimation. In IEEE INFOCOM, 2004.
- [8] E. Lua, T. Griffin, M. Pias, H. Zheng, and J. Crowcroft. On the Accuracy of Embeddings for Internet Coordinate Systems. In ACM Internet Measurements Conference, 2005.
- [9] M. Kaafar, L. Mathy, T. Turetti and W. Dabbous. Virtual Networks under Attack: Disrupting Internet Coordinate Systems. In CoNEXT Conference, 2006.
- [10] M. Kaafar, L. Mathy, C. Barakat, K. Salamatian, T. Turetti and W. Dabbous. Securing Internet Coordinate System: Embedding Phase. In ACM SIGCOMM, 2007.
- [11] J. Ledlie, P. Gardner and M. Seltzer. Network Coordinates in the Wild. In NSDI, 2007.
- [12] D. Stutzbach and R. Rejaie. Understanding Churn in P2P Networks. In ACM Internet Measurements Conference (IMC), 2006.
- [13] M. Crovella and B. Krishnamurthy. Internet Measurement: Infrastructure, Traffic and Applications. Published by Wiley, 2006.
- [14] N. Spring, R. Mahajan and D. Wetherall. Measuring ISP Topologies with Rocketfuel. In ACM SIGCOMM, 2002.
- [15] S. Rewaskar and J. Kaur. Testing the Scalability of Overlay Routing Infrastructures. In Passive and Active Measurements (PAM), 2004.
- [16] A. Parker. The true picture of P2P filesharing. Online: <http://www.cachelogic.com>, July 2004.
- [17] Bindal, et.al. Improving Traffic Locality in BitTorrent via Biased Neighbour Selection. In IEEE ICDCS, 2006.
- [18] S. Ratnasamy, M. Handley, R. Karp and S. Shenker. Topologically aware Overlay Construction and Server Selection. In IEEE INFOCOM, 2002.
- [19] B. Wong, I. Stoyanov and E. Sirer. Octant: A Comprehensive Framework for the Geolocalization of Internet Hosts. In NSDI, 2007.
- [20] S. Halabi. Internet Routing Architectures. Published by Cisco Press, 2000.