ACCEPTED FROM OPEN CALL

A Survey of Research on the Application-Layer Traffic Optimization Problem and the Need for Layer Cooperation

Vijay K. Gurbani, Volker Hilt, Ivica Rimac, and Marco Tomsu, Bell Laboratories, Alcatel-Lucent
Enrico Marocco, Telecom Italia Lab

ABSTRACT

A significant part of Internet traffic today is generated by peer-to-peer applications, used traditionally for file sharing, and more recently for real-time communications and live media streaming. Such applications discover a route to each other through an overlay network with little knowledge of the underlying network topology. As a result, they may choose peers based on information deduced from empirical measurements, which can lead to suboptimal choices. We refer to this as the application-layer traffic optimization (ALTO) problem and present a survey of existing literature. We summarize and compare existing approaches, identify open research issues, and state the need for layer cooperation as a solution to the ALTO problem.

INTRODUCTION AND PROBLEM STATEMENT

A significant part of today's Internet traffic is generated by peer-to-peer (P2P) applications, used originally for file sharing, and more recently for real-time multimedia communications and live media streaming. P2P applications cause between 40 and 85 percent of all Internet traffic and thus pose one of the most serious challenges to the Internet infrastructure.

P2P systems ensure that popular content is replicated at multiple instances in the overlay. But, perhaps ironically, a peer searching for that content may ignore the topology of the network and instead select among available instances based on information it deduces from empirical measurements, which in some particular situations may lead to suboptimal choices. For example, a shorter round-trip time (RTT) estimation is not indicative of the bandwidth and reliability of the underlying links, which have more of an influence than delay on large file transfer P2P applications. Thus, it would appear that P2P networks with their application layer routing strategies based on overlay topologies are in direct competition with Internet routing and topology.

Most distributed hash tables (DHTs) — the data structure that imposes a specific ordering for P2P overlays — use greedy forwarding algorithms to reach their destination, making locally optimal decisions that may not turn to be globally optimized [1]. This naturally leads to what we refer to as the application-layer traffic optimization (ALTO) problem: how to provide the right type of network information to the requesting peer to enable it to perform better than random initial peer selection.

One way to solve the ALTO problem is to build distributed application-layer services for location and path selection [2–7] in order to enable peers to estimate their position in the network and efficiently select their neighbors. Similar solutions have been embedded into P2P applications such as Azureus.[1] A slightly different approach is to have the Internet service provider (ISP) take a proactive role in the routing of P2P application traffic; the means by which this can be achieved have been proposed [8–10]. There is an intrinsic struggle between the layers — P2P overlay and network underlay — when performing the same service (routing); however, there are strategies to mitigate this dichotomy [11]. Our position in this article is that solutions to the ALTO problem will be best achieved by enabling communications between the P2P application layer and the network layer.

The rest of this article is structured as follows. We survey the existing literature on topology estimation and layer interactions. We make a case for our position on the need for layer cooperation. We detail the open research issues that need to be addressed for layer cooperation. Finally, we conclude the article.

SURVEY OF EXISTING LITERATURE

The interaction between application-layer overlays and the underlying networks continues to be a rich area for investigation. The available literature in

[1] *http://www.azureus.com*

Network coordinate systems require the embedding of the Internet topology into a coordinate system. This is not always possible without errors, which impacts the accuracy of distance estimations.

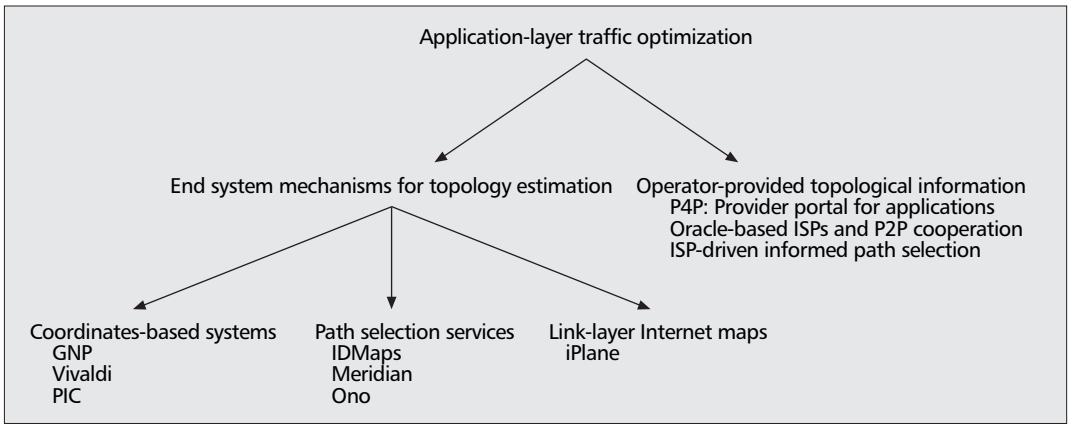

Figure 1. *Taxonomy of solutions for the application-layer traffic optimization problem.*

this field can be divided into two categories (Fig. 1): application-layer techniques to estimate topology and techniques where the application-layer works closely with some other layer (e.g., routing) or some entity that has more of a global networking view than the application does in isolation.

End System Mechanisms for Topology Estimation

Estimating network topology information on the application layer has been an area of active research. Early systems used triangulation techniques to bound the distance between two hosts using a common landmark host. In such a technique, given a cost function C, a set of vertices $V = \{a, b, c\}$ and their corresponding edges, the triangle inequality holds if vertices in V, $C(a, c) \leq C(a, b) + C(b, c)$. The cost function C could be expressed in terms of desirable metrics such as bandwidth or latency.

We note that the techniques presented in this section are only representative of the sizable research in this area. Space does not permit us to enumerate an exhaustive list; rather, we have chosen certain techniques because they represent an advance in the area that further led to other derivative works.

Francis *et al.* proposed IDMaps [2], a system where one or more special hosts called tracers are deployed near an autonomous system. The distance between hosts A and B is estimated as the cumulative distance between A and its nearest tracer T_1, plus the distance between B and its nearest tracer T_2, plus the shortest distance from T_1 to T_2. To aid in scalability beyond that provided by the client-server design of IDMaps, Ng *et al.* proposed a P2P-based global network positioning (GNP) architecture [3]. GNP was a network coordinate system based on absolute coordinates computed from modeling the Internet as a geometric space. It proposed a two-part architecture: in the first part, a small set of finite distributed hosts called landmarks compute their own coordinates in a fixed geometric space. In the second part, a host wishing to participate computes its own coordinates relative to those of the landmark hosts. Thus, armed with the computed coordinates, hosts can then determine interhost distances as soon as they discover each other.

Both IDMaps and GNP require fixed network infrastructure support in the form of tracers or landmark hosts; this often introduces a single point of failure and inhibits scalability. To combat this, new techniques were developed that embedded the network topology in a low-dimensional coordinate space to enable network distance estimation through vector analysis. Costa *et al.* introduced Practical Internet Coordinates (PIC) [5]. While PIC uses the notion of landmark hosts, it does not require explicit network support to designate specific landmark hosts. Any node whose coordinates have been computed could act as a landmark host. When a node n joins the system, it probes the network distance to some landmark hosts. Then it obtains the coordinates of each landmark host and computes its own coordinates relative to the landmark host, subject to the constraint of minimizing the error in the predicted distance and computed distance.

Like PIC, Vivaldi [4] proposed a fully distributed network coordinate system without any distinguished hosts. Whenever a node A communicates with another node B, it measures the RTT to that node and learns that node's current coordinates. A subsequently adjusts its coordinates such that it is closer to or further from B by computing new coordinates that minimize the squared error. A Vivaldi node is thus constantly adjusting its position based on a simulation of interconnected mass springs. Vivaldi is now used in the popular P2P application Azureus, and studies indicate that it scales well to very large networks [12].

Network coordinate systems require the embedding of the Internet topology into a coordinate system. This is not always possible without errors, which impacts the accuracy of distance estimations. In particular, it has proven to be difficult to embed the triangular inequalities found in Internet path distances [12]. Thus, Meridian [6] abandons the generality of network coordinate systems and provides specific distance evaluation services. In Meridian each node keeps track of a small fixed number of neighbors and organizes them in concentric rings, ordered by distance from the node. Meridian locates the closest node by performing a multihop search where each hop exponentially reduces the distance to the target. Although less general than virtual coordinates, Meridian incurs significantly less error for closest node discovery.

The Ono project [13] takes a different approach and uses network measurements from a content distribution network (CDN) like Akamai to find nearby peers. Used as a plug-in to the Azureus BitTorrent client, Ono provides 31 percent average download rate improvement.

Table 1 summarizes the application layer topology estimation techniques. The salient performance metric is relative error. While all approaches define this metric a bit differently, it can be generalized as how close a predicted distance comes to the corresponding measured distance. A value of zero implies perfect prediction, and a value of 1 implies that the predicted distance is in error by a factor of two. PIC, Vivaldi, and Meridian compare their results with that of GNP, while GNP itself compares its results with a precursor technique, IDMaps. Because each of the techniques uses a different Internet topology and a varying number of landmarks and dimensions to interpret the data set, it is impossible to normalize the relative error across all techniques uniformly. Thus, we present the relative error data in pairs, as reported in the literature describing the specific technique. Readers are urged to compare the relative error performance in each column on its own and not draw any conclusions by comparing the data across columns.

Most of the work on estimating topology information focuses on predicting network distance in terms of latency and does not provide estimates for other metrics such as throughput or packet loss rate. However, for many P2P applications latency is not the most important performance metric, and these applications could benefit from a richer information plane. Sophisticated methods of active network probing and passive traffic monitoring are generally very powerful, and can generate network statistics indirectly related to performance measures of interest, such as delay and loss rate on link-level granularity. Extraction of these hidden attributes can be achieved by applying statistical inference techniques developed in the field of inferential network monitoring or network tomography subsequent to sampling of the network state. Thus, network tomography enables the extraction of a richer set of topology information, but at the same time inherently increasing complexity of a potential information plane and introducing estimation errors. For both active and passive methods, statistical models for the measurement process need to be developed, and the spatial and temporal dependence of the measurements should be assessed. Moreover, measurement methodology and statistical inference strategy must be considered jointly. For a deeper discussion of network tomography and recent developments in the field, we refer the reader to [14].

One system providing such a service is iPlane [7], which aims at creating an annotated atlas of the Internet that contains information about latency, bandwidth, capacity, and loss rate. To determine features of the Internet topology, iPlane bridges and builds on different ideas, such as active probing based on packet dispersion techniques to infer available bandwidth along path segments. These ideas are drawn from different fields, including network measurement and network tomography.

Landmark support				
IDMaps	GNP	PIC	Vivaldi	Meridian
Yes	Yes	No	No	No
90th percentile relative error Results in terms of number of (D)imensions and (L)andmarks				
PIC[a] GNP	IDMaps[b] GNP	Vivaldi GNP	Meridian GNP	
8D/16L	7D/15L	2D/32L	8D/15L	
GNP: 0.37	GNP: 0.50	GNP: 0.65	GNP: 1.18	
PIC: 0.38	IDMaps: 0.97	Vivaldi: 0.65	Meridian: 0.78	

[a] Using results from the hybrid strategy for PIC

[b] Does not use dimensions or landmarks

Table 1. *Summary of end system topology estimation techniques.*

Operator-Provided Topological Information

Instead of estimating topology information by end systems through distributed measurements, this information could be provided by the entities running the physical networks — usually ISPs or network operators. In fact, through the constantly updated databases operators use for traffic engineering and network-layer path computational elements (PCE) used for routing, operators already have full knowledge of the topology and general health of the networks they administer. In addition, operators directly know about the business agreements and routing policies that are used to shape traffic in the physical network. In order to avoid congestion on critical links and conflicts between application-level traffic optimization and traffic engineering, operators may be interested in helping applications optimize the traffic they generate.

The remainder of this section briefly describes three recently proposed solutions that follow such an approach to address the ALTO problem.

P4P Architecture — The architecture proposed by Xie *et al.* [8] has been adopted by the DCIA P4P Working Group ,[2] an open group established by ISPs, P2P software distributors, and technology researchers with the dual goal of defining mechanisms to accelerate content distribution and optimize utilization of network resources.

The main role in the P4P architecture is played by servers called iTrackers, deployed by network providers and accessed by P2P applications (or, in general, by elements of the P2P system) in order to make optimal decisions when selecting a peer to connect. An iTracker may offer three interfaces:

- Info: Allows P2P elements (e.g., peers or trackers) to get opaque information associated with an IP address. Such information is kept opaque to hide the actual network topology, but can be used to compute the network distance between IP addresses
- Policy: Allows P2P elements to obtain poli-

[2] *http://www.dcia.info/activities/#P4P*

When using network coordinates to estimate topology information the underlying assumption is that distance in terms of latency determines performance. However, for file sharing and content distribution applications there is more to performance than just the network latency between nodes.

cies and guidelines of the network, which specify how a network provider would like its networks to be utilized at a high level, regardless of P2P applications
- Capability: Allows P2P elements to request network providers' capabilities

The P4P architecture is under evaluation through simulations and experiments on the PlanetLab distributed testbed, and field tests with real users. Initial simulations and PlanetLab experimental results indicate that improvements in BitTorrent download completion time and link utilization in the range of 50–70 percent are possible. Results observed in field tests conducted with a modified version of the software used by the Pando content delivery network show improvements in download rate of 23 percent and a significant drop in data delivery average hop count (from 5.5 to 0.89) in certain scenarios [8].

Oracle-Based ISP-P2P Collaboration — In the general solution proposed by Aggarwal *et al.* [9], network providers host servers called oracles that help P2P users choose optimal neighbors.

The mechanism is fairly simple: a P2P user sends the list of potential peers to the oracle hosted by its ISP, which ranks such a list based on its local policies. For instance, the ISP can prefer peers within its network to prevent traffic from leaving its network; furthermore, it can pick higher-bandwidth links or peers that are geographically closer. Once the application has obtained an ordered list, it is responsible for establishing connections with a number of peers it can individually choose, but it has enough information to perform an optimal choice.

Such a solution has been evaluated with simulations and experiments run on the PlanetLab testbed, and the results show both improvements in content download time and a reduction of overall P2P traffic, even when only a subset of the applications actually query the oracle to make their decisions.

ISP-Driven Informed Path Selection (IDIPS) Service — The solution proposed by Saucez *et al.* [10] is essentially a modified version of the oracle-based approach described above, and is intended to provide a network-layer service for finding best source and destination addresses when establishing a connection between two endpoints in multihomed environments (which are common in IPv6 networking). Peer selection optimization in P2P systems — the ALTO problem in today's Internet — can be addressed by the IDIPS solution as a specific subcase where the options for the destination address consist of all the peers sharing a desired resource, while the choice of the source address is fixed. An evaluation performed on IDIPS shows that costs for both providing and accessing the service are negligible.

The Case for Layer Cooperation as a Solution to the ALTO Problem

The application-level techniques described above provide tools for P2P applications to estimate parameters of the underlying network topology. Although these techniques can improve application performance, there are fundamental limitations on what can be achieved by operating only on the application level.

Topology estimation techniques use abstractions of the network topology, which often hide features that would be of interest to the application. Network coordinate systems, for example, are unable to detect overlay paths shorter than the direct path in the Internet topology. However, these paths frequently exist in the Internet [12]. Similarly, application-level techniques may not accurately estimate topologies with multipath routing.

When using network coordinates to estimate topology information, the underlying assumption is that distance in terms of latency determines performance. However, for file sharing and content distribution applications there is more to performance than just the network latency between nodes. The utility of a long-lived data transfer is determined by the throughput of the underlying TCP protocol, which depends on the round-trip time (RTT) as well as the loss rate experienced on the corresponding path. Hence, these applications benefit from a richer set of topology information that goes beyond latency, including loss rate, capacity, and available bandwidth.

Some of the topology estimation techniques used by peer-to-peer applications need time to converge to a result. For example, current BitTorrent clients implement local passive traffic measurements and a tit-for-tat bandwidth reciprocity mechanism to optimize peering selection at a local level. Peers eventually settle on a set of neighbors that maximizes their download rate, but because peers cannot reason about the value of neighbors without actively exchanging data with them, and the number of concurrent data transfers is limited (typically to 5–7), convergence is delayed and can easily be suboptimal.

Skype's P2P VoIP application chooses a relay node in cases where two peers are behind NATs and cannot connect directly. Ren *et al.* [15] discovered that the relay selection mechanism of Skype:
- Is not able to discover the best possible relay nodes in terms of minimum RTT
- Requires a long setup and stabilization time, which degrades the end-user experience
- Creates a non-negligible amount of overhead traffic due to probing a large number of nodes

They further showed that the quality of the relay paths could be improved when the underlying network AS topology is considered.

Some features of the network topology are hard to infer through application-level techniques, and it may not be possible to infer them at all. Examples of such features are service provider policies and preferences such as the state and cost associated with interdomain peering and transit links. Another example is the traffic engineering policy of a service provider, which may counteract the routing objective of the overlay network, leading to poor overall performance [11].

Finally, application-level techniques often require applications to perform measurements on the topology. These measurements create traffic overhead, in particular, if measurements are performed individually by all applications interested in estimating topology.

Given these problems of application-level topology estimation techniques, we argue that a better solution involves layer cooperation.

Open Research Issues

A low-cost approach to encourage traffic optimization through locality is to use caching. While useful, caching has at least two drawbacks: first, a caching solution depends on the protocol (HTTP, BitTorrent) used to access the resource, thus relegating it as a solution for the most popular applications. Second, copyright issues can prohibit caching the most popular resources. However, solutions based on topology estimation and layer cooperation in general do not interfere with caching. On the contrary, if the peer selection service used by applications is aware of the presence of caches, it can give them higher priorities in its responses and thus achieve greater optimization.

Beyond caching, we believe that there are sizable open research issues to address in an infrastructure-based approach to traffic optimization. The following is not an exhaustive list, but a representative sample of the pertinent issues.

A resilient protocol. For layer cooperation, a solution for the ALTO problem will require a resilient protocol. The exact nature of such a protocol is an open research topic; it could be a simple request-response protocol based on HTTP that leaves the bulk of the routing decision on the P2P host, which simply consults a topological database provided by the operator. Or the protocol could be much more fine-grained, based on a common ontology such that the request contains certain preferences the operator takes into account to provide an optimized routing decision to the P2P host. Regardless, such a protocol should adequately express the preferences of a P2P host to the network operator without disclosing information the host may consider private, providing the network operator with enough primitives such that the response sent back to the host does not reveal undue topological information of the operator's network. Ideally, the protocol should be composable such that the response from a previous request is used as an input vector to a subsequent request. A canonical example of such a composite service would be a peer receiving routing instructions from a localization service and sending them to a cost service that aims to minimize the monetary costs associated with fetching the contents by perhaps delaying file transfer until a suitable time in the near future. Another important consideration for such a protocol is that it be scalable; protocol design considerations such as idempotence, statelessness, and use of connection-oriented or connectionless transports are all issues that impact scalability.

Coordinate estimation or path latencies? Despite the many solutions that have been proposed for providing applications with topology information in a fully distributed manner, there is currently an ongoing debate in the research community on whether such solutions should focus on estimating nodes' coordinates or path latencies. Such a debate has recently been fed by studies showing that the triangle inequality on which coordinate systems are based is often proved false in the Internet [12]. Proposed systems following both approaches — in particular, Vivaldi [4] and PIC [5] following the former, Meridian [6] and iPlane [7] the latter — have been simulated, implemented, and studied in real-world trials, each showing different points of strength and weakness. Concentrated work will be needed to determine which of the two will be conducive to the ALTO problem.

Malicious nodes. Another open issue common in most distributed environments consisting of a large number of peers is resistance against malicious nodes. Security mechanisms to identify misbehavior are based on triangle inequality checks [5], which tend to fail and thus return false positives in the presence of measurement inaccuracies induced, for example, by traffic fluctuations that occur quite often in large networks [12]. Beyond the issue of using triangle inequality checks, authoritatively authenticating the identity of an oracle and protecting an oracle from attacks are also important. Exploration of existing techniques, such as public key infrastructure or identity-based encryption for authenticating identity and the use of secure multiparty computation techniques to prevent an oracle from collusion attacks, need to be studied for judicious use in ALTO solutions.

Information integrity. Similarly, even in controlled architectures deployed by network operators where system elements may be authenticated [8–10], it is still possible that the information returned to applications is deliberately altered; for example, assigning higher priority to cheap (monetary-wise) links instead of neutrally applying proximity criteria. What are the effects of such deliberate alterations if multiple peers collude to determine a different route to the target, one that is not provided by an oracle? Similarly, what are the consequences if an oracle targets a particular node in another AS by redirecting an inordinate number of querying peers to it, essentially causing a distributed denial-of-service (DDoS) attack on the node? Furthermore, does an oracle broadcast or multicast a response to a query? If so, techniques to protect the confidentiality of the multicast stream will need to be investigated to thwart "free riding" peers.

Simulate or build? Much debate in the P2P research community clusters around the simulate or build question. Undoubtedly, it is hard to foresee how proposed systems will perform in the Internet. Simulations and testbed emulations are in most cases the only options available for benchmarking the performance of a system. However, these have often proven inadequate; in at least one particular case [12], they have only provided rough optimistic approximations of what would be measured in the real world. Even using near-realistic testbeds such as PlanetLab does not suffice for certain aspects of quantifying P2P traffic: more often, these testbeds do not take in account the user component, which is crucial for file-sharing P2P systems. After all, a P2P system depends on the choices and interests of its users to fetch, store, and disseminate content, and it is hard to simulate a sizable user population with varying tastes to authoritatively observe the behavior of a P2P network. New techniques in simulation or testbed usage need to be investigated.

Richness of topological information. Many systems already use RTT to account for delay when establishing connections with peers (e.g., CAN, Bamboo). An operator can provide not only the delay metric, but other metrics as well that the peer cannot

Solutions based on topology estimation and layer cooperation in general do not interfere with caching. On the contrary, if the peer selection service used by applications is aware of the presence of caches, it can give them higher priorities in its responses and thus achieve greater optimization.

Due to the impact of P2P traffic on the network, the IETF has started to investigate the standardization of a protocol between applications and network elements (an ALTO server, for instance) to aid in preferential peer selection while taking in account the network topology.

figure out on its own. These metrics may include the characteristics of the access links to other peers, bandwidth available to peers (based on the operator's engineering of its network), network policies, and preferences such as state and cost associated with intradomain peering links. Exactly what kinds of metrics an operator can provide to stabilize the network throughput also needs to be investigated.

Hybrid solutions. It is conceivable that P2P users may not be comfortable with operator intervention to provide topology information. To eliminate this intervention, alternative schemes to estimate topological distance can be used. For instance, Ono uses client redirections generated by Akamai CDN servers as an approximation for estimating distance to peers; Vivaldi, GNP, and PIC use synthetic coordinate systems. A neutral third party can make available a hybrid layer cooperation service — without the active participation of the ISP — that uses alternative techniques discussed earlier to create a topological map. This map can be subsequently used by a subset of users who may not trust the ISP.

Conclusion

We have argued that it is beneficial for a solution to the ALTO problem to involve cross-layer cooperation, allowing communications between applications and network elements aware of the underlying network topology. In particular, such a solution should specify the following:

- A lookup mechanism to be used by applications to discover the appropriate network elements to query in order to obtain topology information they need for ALTO
- A protocol to be used in communications between applications and those network elements

Due to the impact of P2P traffic on the network, the Internet Engineering Task Force (IETF) has started to investigate the standardization of a protocol between applications and network elements (e.g., an ALTO server) to aid in preferential peer selection while taking into account the network topology. A workshop sponsored in May 2008 on this topic lead to a BoF session at the 72nd meeting (the BoF agenda and presentations are archived at http://www.ietf.org/proceedings/08jul/agenda/alto.html) and subsequently to the creation of the ALTO working group at the 73rd meeting (http://www.ietf.org/html.charters/alto-charter.html).

Acknowledgments

Stefano Previdi provided insight into PCEs with respect to ALTO. Henning Schulzrinne first described the composite service to minimize operator costs mentioned earlier. Finally, we thank profusely the anonymous reviewers of this article for their insightful feedback.

References

[1] K. Gummadi *et al.*, "The Impact of DHT Routing Geometry on Resilience and Proximity," *Proc. ACM SIGCOMM*, Aug. 2003.
[2] P. Francis *et al.*, "A Global Internet Host Distance Estimation Service," *IEEE/ACM Trans. Net.*, Oct. 2001.
[3] T. S. E. Ng, and H. Zhang. Predicting Internet Network Distance with Coordinates-Based Approaches," *Proc. IEEE INFOCOM*, June 2002.
[4] F. Dabek *et al.*, "Vivaldi: A Decentralized Network Coordinate System," *Proc. ACM SIGCOMM*, Aug. 2004.
[5] M. Costa *et al.*, "PIC: Practical Internet Coordinates for Distance Estimation," *Int'l. Conf. Distrib. Sys.*, Tokyo, Japan, Mar. 2004.
[6] B. Wong, A. Slivkins, and E. G. Sirer, "Meridian: A Lightweight Network Location Service Without Virtual Coordinates," *Proc. ACM SIGCOMM*, Aug. 2005.
[7] H. V. Madhyastha *et al.*, "iPlane: An Information Plane for Distributed Services," *Proc. OSDI*, Nov. 2006.
[8] H. Xie *et al.*, "P4P: Provider Portal for Applications," *ACM SIGCOMM Comp. Commun. Rev.*, vol. 38, no. 4, Oct. 2008, pp. 351–62.
[9] V. Aggarwal, A. Feldmann, and C. Scheidler, "Can ISPs and P2P Systems Co-Operate for Improved Performance?" *ACM SIGCOMM Comp. Commun. Rev.*, vol. 37, no. 3, July 2007, pp. 29–40.
[10] D. Saucez, B. Donnet, and O. Bonaventure, "Implementation and Preliminary Evaluation of an ISP-Driven Informed Path Selection," *Proc. ACM CoNEXT*, Dec. 2007.
[11] S. Seetharaman *et al.*, "Preemptive Strategies to Improve Routing Performance of Native and Overlay Layers," *Proc. IEEE INFOCOM,* May 2007.
[12] J. Ledlie, P. Gardner, and M. Seltzer, "Network Coordinates in the Wild," *Proc. NSDI*, Cambridge, MA, Apr. 2007.
[13] D. R. Choffnes and F. E. Bustamante, "Taming the Torrent: A Practical Approach to Reducing Cross-ISP Traffic in P2P Systems," *Proc. ACM SIGCOMM '08*, vol. 38, no. 4, Oct. 2008, pp. 363–74.
[14] R. Castro *et al.*, "Network Tomography: Recent Developments," *Statistical Sci.*, vol. 19, no. 3, 2004, pp. 499–517.
[15] S. Ren, L. Guo, and X. Zhang, "ASAP: An AS-Aware Peer-Relay Protocol for High Quality VoIP," *Proc. IEEE Int'l. Conf. Distrib. Comp. Sys.*, 2006, pp. 70–79.

Biographies

Vijay Gurbani (vkg@bell-labs.com) works for Bell Laboratories, Alcatel-Lucent. He holds B.Sc. and M.Sc. degrees in computer science from Bradley University, and a Ph.D. in computer science from Illinois Institute of Technology. He currently works on the application of P2P techniques in different domains. He is the author of six IETF RFCs, 13 patents (granted or pending), three books, and numerous journal and conference proceedings. He is a senior member of the ACM and a member of the IEEE Computer Society.

Volker Hilt (volkerh@bell-labs.com) works for Bell Labs/Alcatel-Lucent where he manages a department in the Services Infrastructure Research Domain. His research interests include P2P technologies, content distribution networks, real-time and group communication, and multimedia applications. He has contributed to several standards in the area of multimedia communication in the IETF and chairs the P2P research group in the Internet Research Task Force (IRTF). He received his Master's degree in computer science and business administration, and his Ph.D. in computer science, both from the University of Mannheim, Germany.

Ivica Rimac (rimac@bell-labs.com) is a member of the Services Infrastructure Research Domain at Bell Laboratories, Holmdel, New Jerseyu. He obtained his Ph.D. and M.S. in electrical engineering and information technology from Darmstadt University of Technology, Germany. He is currently working on technologies for content dissemination in fixed and mobile communication networks leveraging network caching and P2P techniques. He has co-authored 10 patents (granted or pending) and published 20+ research papers.

Marco Tomsu (marco.tomsu@alcatel-lucent.com) is a project leader within Alcatel-Lucent Bell Labs Service Infrastructures Research Domain. Currently based in Stuttgart, Germany, he received a Dipl.-Ing. degree in communications engineering from the University of Karlsruhe. He is a member of the Alcatel-Lucent Technical Academy. He holds patents and has co-authored papers in several technical areas. He is contributing to IETF standardization and served as guest editor of the *Bell Labs Technical Journal* issue on content networking.

Enrico Marocco (enrico.marocco@telecomitalia.it) is a research engineer in the Service Innovation Area at Telecom Italia. He was involved in the design and deployment of the first VoIP network at Telecom Italia, with special focus on standard compliancy. He is active in IETF, where he serves as chair of the ALTO Working Group. He is also involved in open source activities and leads the SIPDHT project. His research interests include P2P communications and content distribution.