Diversifying the Internet

Jonathan S. Turner[1]
Jon.Turner@wustl.edu

David E. Taylor[1,2]
david@exegy.com

[1]Applied Research Laboratory
Washington University in Saint Louis

[2]Exegy Inc.

Abstract—**The Internet has fallen victim to its own stunning success. The interplay of the end-to-end design of IP and the vested interests of competing stakeholders has led to its growing ossification. Alterations to the Internet architecture that address its fundamental deficiencies or enable new services have been restricted to incremental changes. The slow pace of this process stifles innovation and the adoption of disruptive technology. A recent call to arms advances a research agenda to confront this impasse through virtualization [1]. In addition to describing a virtual testbed for the evaluation of new network architectures, it poses a question about the long-term role of virtualization in the Internet. The architectural "purist" views virtualization as a tool for architecture evaluation and the periodic deployment of successive, singular Internet architectures. In this paper, we advance the "pluralist" view that seeks to make virtualization an architectural attribute of the Internet. By enabling a plurality of diverse network architectures to coexist on a shared physical substrate, virtualization mitigates the ossifying forces at work in the current Internet and enables continual introduction of innovative network technologies. Such a diversified Internet would allow existing architectural deficiencies to be holistically addressed as well as enable the introduction of new architectures supporting new types of applications and services. We provide a detailed exposition of the diversified Internet concept, explain how it can address the problem of network ossification and discuss some of the technical challenges that must be met to turn the vision into reality.**

I. Network ossification

In a relatively short period of time, the Internet has become critical infrastructure for global commerce, media, and defense. Like many successful technologies, the Internet is suffering the adverse effects of inertia. The significant capital investment and competing interests of its major stakeholders creates a barrier to the introduction of disruptive technologies. Furthermore, the end-to-end design of IP requires global agreement and coordination to deploy changes. Working in concert, these two factors prevent existing problems from being holistically addressed and stifle innovation. In the current environment, the incremental deployment of even the most basic and necessary changes such as IPv6 is painfully slow [2].

While the Internet's success speaks to the utility of its architecture for a wide range of applications, a best-effort packet delivery service is not best for all purposes. The ever-expanding scope and scale of the Internet's use has also exposed a number of fundamental deficiencies in the current architecture. Security, routing stability and control, and quality of service guarantees are a few of the most significant [3-5]. For over ten years, applications such as global video conferencing, telephony, and broadcast television have been touted as promising "next-generation" applications. It was widely believed that this class of applications justified the tremendous investment in dark fiber and would spur the next round of vigorous innovation and deployment [6]. This promise remains largely unrealized due to the inability of the current architecture to support these applications, the inability to change the architecture, and the prohibitively high cost of deploying a custom global network.

The ossification of the Internet architecture and its inability to support many types of applications impose a significant barrier to innovation. For the research community, the ability to perform large-scale experimentation that seeks to affect the Internet's operation is almost entirely limited to overlay networks. It can bring innovation to the network layer only through more efficient implementations of existing mechanisms. While there is opportunity to deploy new higher layer protocols and new link and physical layer technologies, the network layer has become untouchable.

II. Enabling diversity

Virtualization has been advanced as a means to break this impasse [1, 7, 8]. By enabling experimentation with new network architectures, a virtual testbed provides immediate value to the research community. We discuss the development of such a testbed in the last section of this paper. While reinvigoration of applied architectural research has clear benefits, it is the introduction of new technologies into the public Internet that can provide the more profound impact on society and which also presents the most significant challenge.

In the context of deployment, we characterize the architectural purist as viewing virtualization as a tool for experimentation and the periodic deployment of each in a possible succession of new Internet architectures. The theory here is that by supporting applications that attract a significant number of users, a successful network architecture operating on a virtual testbed can apply pressure on the current Internet stakeholders to adopt a new architecture. In this environment, experimentation may be continual, but adoption of a new network architecture is relatively rare. We seek to advance the pluralist view that virtualization should become a fundamental attribute of the Internet so that the problem of network ossification can be solved, once and for all.

In crafting a network, architects must invariably choose from among numerous parameters. One of the first considerations

This work supported by the National Science Foundation, CNS-0435173.

is often scope. Networks may be application-specific, general-purpose, or a hybrid that supports a certain class of applications. Networks may provide best-effort service or multiple flavors of guaranteed service. Other considerations include the type and level of security, control granularity and interfaces, and network scale. These architectural parameters then drive the design of the data and control planes. The architectural purist requires that a network architecture be general-purpose. It must provide a suitable platform for the set of all existing applications, while attempting to support the needs of future applications [9]. If the architecture is insufficient for a given application, then developers may attempt to craft a suitable overlay with sub-optimal properties (security, robustness, or performance) or simply wait until a new, more suitable architecture is adopted. The purist view also requires that each network node provide all of the shared services, regardless of the needs of the application traffic traversing the node. More importantly, we believe that the purist view does not sufficiently address the ossifying forces currently at work in the Internet. If the "narrow waist" of the Internet is a single end-to-end packet delivery service, then modification of that packet delivery service still requires universal agreement and coordination amongst the competing stakeholders. In this environment, virtualization (i.e. a virtual testbed) does not possess significantly more leverage than previous testbeds that have been unable to successfully impact the Internet.

The pluralist approach provides a means to permanently guard against these ossifying forces while providing additional design freedom for network architects [10]. In the remainder of this paper, we advance this view by describing a diversified Internet that supports a plurality of network architectures coexisting on a shared physical substrate network. We refer to the diverse networks comprising this diversified Internet as meta-networks to distinguish the concept from overlay [8] and underlay [7] approaches, and to avoid the much overloaded term, "virtual network". A diversified Internet that supports multiple meta-networks allows new network architectures to be introduced alongside incumbent architectures and given the opportunity to succeed (or fail) on their own merits. By allowing end users to opt-in to a variety of meta-networks, network architectures are exposed to the same market forces as other technologies. We argue that exposing network technologies to such competitive forces is the only sure way to enable continuing innovation and change.

III. Meta-networks

Meta-networks enable fundamentally different end-to-end packet delivery mechanisms by allowing each network to specify packet formats, addressing methods, packet forwarding, routing protocols, etc. Meta-networks are implemented on top of a substrate comprising the physical resources used by the various interlays. In this model, the "narrow waist" of the Internet architecture is a thin network provisioning layer that provides automated mechanisms for allocating the substrate's resources to different meta-networks. The role of the provisioning layer is to allow network providers to automatically deploy, configure and operate meta-networks. In general, the provisioning layer does not restrict the kind of new services that may be deployed.

We define an meta-network to be a network of meta-links joining meta-routers to each other and to the meta-network-specific protocol stacks in end systems. Once provisioned, the meta-links and routers "look" just like physical network resources. It is up to the meta-networks (not the substrate) to implement end-to-end packet delivery by selecting packet formats, forwarding mechanisms, control and management planes, etc. This freedom allows network architects to construct networks that best suit the intended application or range of applications that will operate over the meta-network. The substrate is concerned only with providing the resources needed by distinct meta-networks and ensuring that the different meta-networks can co-exist without interference. Our bias is to include as little as possible in the substrate. While this may mean that different meta-networks duplicate certain common services, it also means that meta-networks with simple needs are not burdened with the cost of supporting some shared service that they don't use. While this potentially increases the effort required to develop new meta-networks, we believe that this can be offset through the use of design tools and the development of modular components that can be used in a variety of different meta-networks.

Applications run on top of meta-networks. Users may opt-in to one or many meta-networks, thus freeing application developers of the constraints imposed by any single network architecture. Developers can target their applications for the meta-network best-suited to their application's requirements. While traditional applications, such as email and the web might well use an IP meta-network, other applications could be designed to target meta-networks with different characteristics. Indeed, an application developer might even choose to develop an application in concert with a new meta-architecture, integrating application and network in ways that have not been possible with conventional network architectures. As an illustration of this, consider a meta-network designed to support distance learning. It would likely include QoS support for audio and video, but might also include specialized multicast mechanisms, with distributed audio bridging (to enable natural, high quality interaction) and audio-triggered video switching, so that the focus of video would track the current speaker. These mechanisms could be incorporated into the meta-routers and coordinated using protocols that operate over an omni-directional multicast tree. Such an meta-network could also include built-in format translators, to enable seamless participation among users with incompatible equipment. While integrating application and network in this way is something that the networking community has made a point to avoid in recent decades, there is no need to avoid it in a diversified network environment.

IV. What's in it for stakeholders?

One lesson we draw from the current ossification of the Internet is that enabling change will require convincing stakeholders that change is in their interest. Fortunately, there is much for current stakeholders to be unhappy about in the Internet as it stands. The best-effort Internet of today is a commodity service that affords network providers limited opportunities to distinguish themselves from their competitors. A diversified Internet

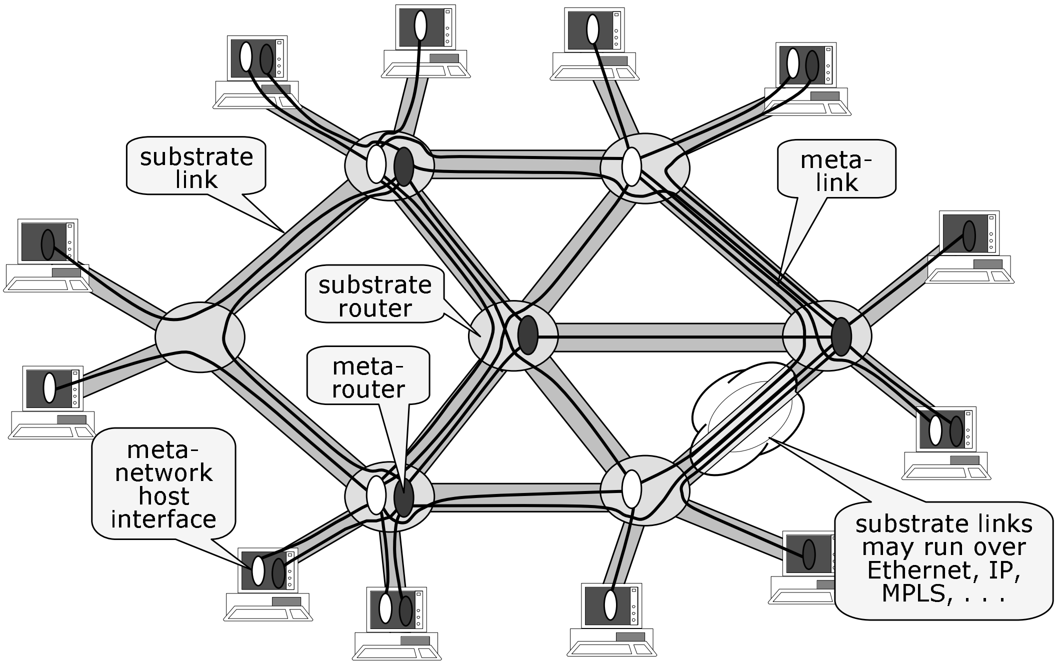
substrate link
meta-link
substrate router
meta-router
meta-network host interface
substrate links may run over Ethernet, IP, MPLS, . . .

Fig. 1. Diversified Internet supporting multiple meta-networks on a shared substrate.

can provide a much richer environment for innovative organizations. This, in turn, can stimulate the development of a wide range of new services and will drive investment in the core infrastructure components needed to deliver those services to end users.

In a diversified internet, the current ISPs become Substrate Network Providers (SNPs), implementing the automated provisioning layer used to configure resources for meta-networks. SNPs can distinguish themselves from their peers through the quality of the resources they provide, the tools they provide to facilitate the development of new meta-networks and the services they provide to operating meta-networks (for example, providing hardware fault tolerance, or integrated end-user billing services).

Meta-networks would be deployed and operated by Meta-Network Providers (MNPs). Meta-networks could span multiple substrate networks, and thus would acquire network resources from multiple SNPs. In this model, MNPs are not required to purchase, deploy, or maintain physical infrastructure. This significantly lowers the barrier to entry and significantly accelerates deployment of new, possibly disruptive, network architectures. SNPs distinguish themselves by the service they provide to meta-network developers and operators. By providing the most attractive platform, SNPs will attract the most innovative meta-providers, helping ensure that the meta-networks that become popular will be carried on their infrastructure.

A diversified Internet also creates new opportunities for network equipment vendors. The systems needed to implement a diversity of meta-routers within a common physical platform will afford vendors with new opportunities to develop unique products. These systems will provide configurable resources that can be used by different meta-routers to deliver their individual services. The emergence of high performance network processors [11] and configurable logic devices [12] provide the core elements needed to build such systems, but the competitiveness of equipment products will rest on how easy vendors make it for meta-network developers to design and configure new types of meta-routers using those resources.

Application developers have a much richer set of choices in a diversified Internet than they have today. They may choose the most suitable meta-network from among those available, or they may design a new overlay network, tailored to their application's specific needs.

V. Resource Allocation in a Diversified Internet

There are several technical challenges that must be addressed to realize the vision of a diversified internet. One of the primary challenges centers on the design of the protocols and associated mechanisms needed to enable the automated provisioning of diverse meta-networks in a global, multi-domain substrate network. The problem to be solved is very different from that addressed by conventional network resource reservation mechanisms. Some of these differences make it easier, others make it more complex. Unlike flow-based reservation where end-to-end resources must be reserved for individual communication sessions, the meta-provisioning system will allocate resources on a relatively coarse-grained and long-term basis. This is because meta-networks will be set up and operated to serve large numbers of users and resources will be reserved to handle the aggregate traffic from these many users over an extended period of time. These differences dramatically reduce the scaling and performance challenges associated with resource provisioning, relative to resource reservation in conventional networks.

On the other hand, the nature of the interaction between meta-networks and a multi-domain substrate is intrinsically more complex than the relationship between a single user and a network (even a multi-domain network) providing a reserved bandwidth flow. An MNP, planning a new meta-network, needs to be able to determine what resources are available within an SNP's network, or could be made available within the MNP's planning horizon. This information must be specific enough to allow the MNP to formulate network designs, possibly involving multiple SNPs. It must then be able to issue a "Request For Bids" that SNPs can respond to, providing further technical details and cost information. This interaction could iterate multiple times as the MNP tries to determine how best to meet its requirements. To enable the automation of this process, SNPs will need to publish certain information in machine-readable form and provide automated mechanisms to respond to RFPs. Once an appropriate configuration has been determined, there must be mechanisms for "closing the deal" and coordinating the actual configuration of meta-links and routers.

VI. Substrate Router Design

A second challenge that must be met in order to realize a diversified Internet concerns the design of substrate routers that can host multiple diverse meta-routers, while remaining competitive, in terms of cost and performance, with conventional routers. In this section we briefly describe a highly flexible architecture for a substrate router and show how it can be used to enable substrate routers with a wide range of performance characteristics and functionality.

Our proposed substrate router architecture is shown in Figure 2. It is built around of pool of *Processing Elements* (PE) which are used to provide the core packet processing functions of the various meta-routers. The system's common substrate comprises a scalable switch fabric and a number of *Line Cards* that terminate the external links. The Line Cards demultiplex packets received on the external links using a Meta-Link Identifier (MLI) carried in the external packet header. The substrate router uses the MLI to resolve the meta-router that the packet is associated with and the specific PE that is to process it. The configuration of the mapping from MLI to meta-router and PE is done at the time an meta-router is configured.

The architecture supports a variety of types of PEs to give meta-router designers flexibility in the implementation of the meta-routers. In particular, the available PE types would include standard general-purpose processor subsystems, a network processor-based PE and a PE comprising a large configurable logic chip, with attached memory resources. We refer to these three PE types as PE/GP, PE/NP and PE/CL respectively. A lower performance meta-router can be built using a PE/GP, while higher performance meta-routers will require the use of PE/NPs or PE/CLs. The availability of multiple PE types allows meta-router designers to choose an implementation option that best suits their needs. In fact, we would expect some meta-routers to use PEs of multiple types (for example, PE/CLs might be used to implement routine packet forwarding in hardware, while a PE/GP is used for higher level control and configuration).

It is worth noting that current-generation network processors provide enough processing resources to deliver approximately 3-5 Gb/s of throughput for moderately complex applications [11], [14], so a substrate router supporting 10 Gb/s line cards can use 2-4 PE/NPs per port, if most packet processing is done using PE/NPs. PE/CLs incorporating advanced FPGAs [12] can deliver throughputs several times greater and can be a better choice for meta-networks that lend themselves to hardware implementation.

In the simplest case, an meta-router consists of a single PE that terminates some number of meta-links. In this case, packets received by the LCs are sent through the switch fabric to the appropriate PE, which processes them, decides how they should be forwarded, then sends the packets back through the switch fabric to the proper outgoing LC and meta-link. In higher performance systems, a single meta-router will comprise multiple PEs. In this case, packets will often be forwarded from one PE to another through the switch fabric. A common pattern will be that packets are first sent to an "input-side" PE, which relays them to an "output-side" PE, which queues them before forwarding them to the outgoing LC and meta-link.

The physical separation of the pool of PEs from the LCs provides great flexibility in the allocation of resources, but this flexiblity does come at some cost. In particular, it requires that each packet must be forwarded through the switch fabric at least two times. The performance impact of this is fairly minimal, since switch fabric delays are generally quite small compared to processing delays, link queueing delays and wide-area propagation delays. However, it does require that the switch fabric capacity be two to three times larger than would be needed by a conventional router terminating the same number of external links. In a well-designed conventional router, the switch fabric typically accounts for perhaps 10% of total system hardware cost, so increasing this cost by a factor of two to three, does not dramatically change the overall system cost. On the other hand, the cost difference is not negligible, making alternatives worth considering.

One natural alternative is to provide flexible packet processing mechanisms at every LC. In this approach, each meta-router terminating a meta-link at a given LC would use a fraction of the processing resources available at that LC. This would enable the meta-routers to forward packets to the appropriate outgoing LC, using just one hop through the switch fabric. While this is attractive, from the standpoint of switch fabric cost, it does require that the processing resources available at each LC be sufficient to handle any processing required by any meta-router that we might want to deploy. It also requires finer-grained resource partitioning than the pool-of-PEs architecture, making it considerably more complex to implement. An intermediate alternative provides some processing capability at each LC, plus a shared pool of PEs. This approach makes a lot of sense if most packets are forwarded by just a few of the meta-routers. These meta-routers can be assigned per LC resources, while the pooled resources are used by meta-routers that handle smaller amounts of traffic. This can significantly reduce switch fabric costs compared to a "pure" pool-of-PEs architecture.

Note that each PE has its own set of resources that are physically separate from those of other PEs. This makes it easy to isolate different meta-routers from one another, through the direct expedient of physical separation. So long as an meta-router has enough traffic to justify allocating an entire PE to it, this provides a simple solution to the isolation problem. In systems hosting many low traffic meta-routers, this may lead

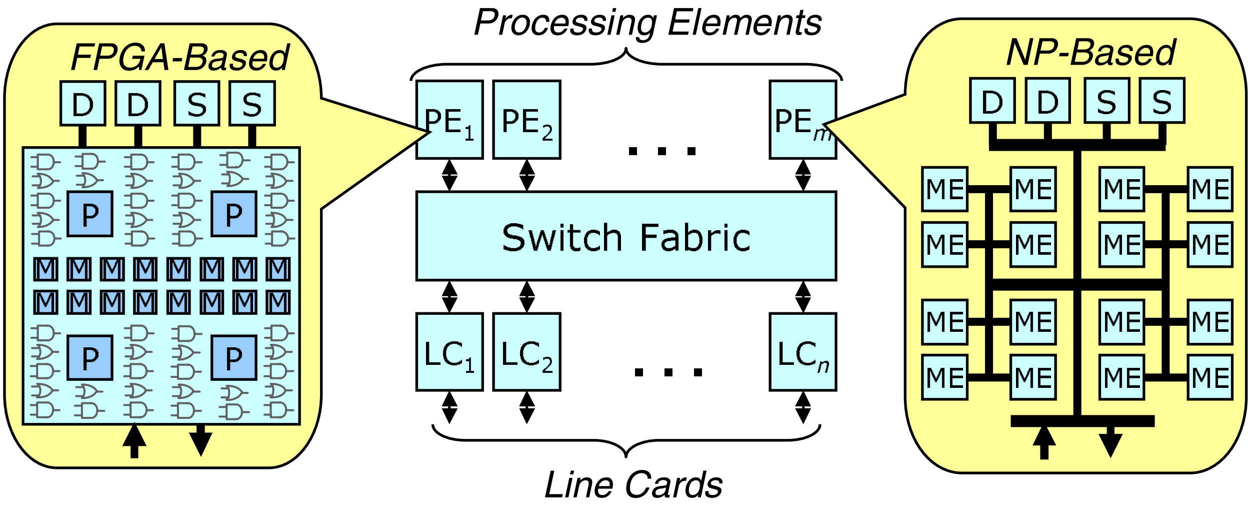

Fig. 2. Substrate router with FPGA based and Network Processor based Processing Elements.

to low resource utilization. In this context, it makes sense to share a single PE among multiple meta-routers. While sharing of PE/GPs can be done using virtual machine methods, there are no comparable methods for sharing PE/NPs and PE/CLs. While the design of such methods is worth further study, we do not pursue the topic here.

There remians the problem of how to isolate the traffic streams associated with different meta-routers within the switch fabric. This is straightforward, in the context of meta-routers that use a single PE, because in this case the traffic consists of point-to-point streams from LC-to-PE or PE-to-LC. Because meta-links have specific bandwidth allocations associated with them, it is straightforward to allocate bandwidth within the switch to the different meta-routers and enforce appropriate limits on their use of bandwidth at the edge of the switch fabric.

Things get more complicated when we consider meta-routers comprising multiple PEs. In this case, we would like to allow the PEs making up an meta-router to freely send packets to other PEs in the same meta-router without imposing pairwise constraints on the amount of traffic one PE can send to another. It is not possible to enforce bandwidth use by an meta-router as the traffic enters the switch fabric in this case, since several different PEs can send traffic to another PE in the same meta-router, leading to congestion in the switch fabric. The challenge for the design of the substrate router is to either prevent such congestion from occurring in the first place or to ensure that the congestion affects only the meta-router causing it. We believe this can be adequately addressed by providing separate queues for different meta-routers within the switch fabric and serving those queues based on the share of the switch fabric resources allocated to each meta-router. Further work is needed to fully evaluate the effectiveness and complexity of this approach.

VII. Getting from Here to There

Perhaps the most immediate challenge facing the concept of a diversified internet is deployment. While we believe that there are great potential benefits to a diversified Internet and that critical stakeholders can be motivated to grasp the opportunity that a diversified Internet affords, we need a strategy for getting from here to there. We argue that this strategy will have to embrace the approaches used in the early deployment of IP. Specifically, we expect it to start as an overlay, progress through a government-supported experimental backbone and shift to commercial operation as network operators recognize the opportunity and as the multi-domain issues become better understood. In some ways, this process has already begun since the ideas we are advancing here are already partially reflected in the PlanetLab testbed [13] which supports overlay meta-networks on top of a shared global infrastructure. The diversified Internet concept carries these ideas further, pushing them into the core of the network, rather than limiting them to the overlay context. We believe that the next step in carrying these ideas forward is the creation of a national testbed that leverages key ideas developed in the PlanetLab context, while refining them to enable the large-scale deployment of diverse meta-networks that can support large numbers of users and serve as a vehicle for new applications [1]. The recent development of the National Lambda Rail (NLR) [16] infrastructure provides an important tool for helping to realize such a testbed. NLR offers a national fiber backbone and is committed to using 50% of its resources for supporting networking research. This makes it realistic to envision a national network with backbone links of more than 10 Gb/s and capable of supporting hundreds of thousands of users. Such a backbone can be linked to PlanetLab nodes in hundreds of colleges and universities throughput the country, using conventional overlay methods or using MPLS tunnels carried on the widely-deployed Internet 2 infrastructure. This, in turn provides access to a potential user population of more than a million. A successful large-scale demonstration of a diversified Internet will stimulate the creation of innovative applications and network services. These, in turn, will attract users and they will attract commercial operators interested in bringing these services to the larger public.

VIII. Summary

We expect the diversification of the Internet to be a long process (as was the development of the Internet itself) and we acknowledge that no one can predict with any confidence how it is likely to play out. However, the problems posed by the Internet's growing ossification make it essential that we find a way to overcome the current impasse and we argue that diversification of the Internet is the most promising approach to both addressing the current problem and avoiding its recurrence in the years to come.

References

[1] L. Peterson, S. Shenker, and J. Turner, "Overcoming the Internet impasse through Virtualization," in *ACM HotNets-III*, 2004.

[2] S. J. Vaughan-Nichols, "We Love IPv6, We Love IPv6 Not," in *Enterprise IT Planet*, 2004.

[3] J. Crowcoft, et al. "QoS's Downfall: At the bottom, or not at all!," in *ACM SIGCOMM Workshop on Future Directions in Network Architecture (FDNA)*, 2003. Karlsruhe, Germany.

[4] N. Feamster, H. Balakrishnan, and J. Rexford, "Some Foundational Problems in Interdomain Routing," in *ACM HotNets-III*, 2004.

[5] V. Sekar, et al. "Toward a Framework for Internet Forensic Analysis", in *ACM HotNets-III*, 2004.

[6] C. Grice, "Qwest expands amid concerns of fiber glut," in *CNET News.com*, 2001.

[7] R. Gold, P. Gunningberg, and C. Tschudin, "A Virtualized Link Layer with Support for Indirection," in *ACM SIGCOMM Workshop on Future Directions in Network Architecture*, 2004. Portland, Oregon.

[8] J. D. Touch, et al. "A Virtual Internet Architecture," in *ACM SIGCOMM Workshop on Future Directions in Network Architecture*, 2003. Karlsruhe, Germany.

[9] D. D. Clark, et al. "Addressing Reality: An Architectural Response to Real-World Demands on the Evolving Internet," in *ACM SIGCOMM Workshop on Future Directions in Network Architecture*, 2003. Karlsruhe, Germany.

[10] J. Crowcoft, et al. "Plutarch: An Argument for Network Pluralism," in *ACM SIGCOMM Workshop on Future Directions in Network Architecture (FDNA)*, 2003. Karlsruhe, Germany.

[11] Intel, "IXP2800 and IXP2850 Network Processors," 2004, Intel Corporation: Datasheet.

[12] Xilinx, "Virtex-II Pro Platform FPGAs: Introduction and Overview," 2003: DS083-1 (v3.0).

[13] Chun, B., et al., "PlanetLab: An Overlay Testbed for Broad-Coverage Services," *ACM Computer Communications Review*, 2003. 33(3).

[14] T. Wolf and M. Franklin, "CommBench - A Telecommunication Benchmark for Network Processors," *IEEE Intl. Symp. on Perf. Analysis of Syst. and Software*, April 2000.

[15] S. Choi, et. al., "Design of a High Performance Dynamically Extensible Router," *Proceedings of the DARPA Active Networks Conference and Exposition*, May 2002.

[16] National LambdaRail, "New Class of National Networking Infrastructure Launched to Support Cutting-Edge Research and Experimentation," September 2003.