Can Overlay Hosting Services Make IP Ossification Irrelevant?

Jon Turner
Washington University
jon.turner@wustl.edu

Nick McKeown
Stanford University
nickm@stanford.edu

There is a sea-change taking place in networking research. In recent years, it has become widely accepted that progress in networking must occur in ways that are incrementally deployable, since the Internet has become too large to allow disruptive changes to its core protocols and services. More recently, growing dissatisfaction with the creeping ossification of the Internet has led to a renewed interest in clean-slate thinking that re-considers many of the basic premises of the Internet architecture, and seeks to create completely new network architectures that correct the deficiencies in the current architecture [AN05].

GENI [GE06] has been proposed as an infrastructure within which new experimental architectures can be deployed and evaluated. GENI will use virtualization techniques to enable multiple experimental networks to co-exist alongside one another, allowing parallel exploration of a diversity of new approaches. Some have argued that virtualization should become a central feature of the next generation Internet, allowing new network architectures to be deployed in the Internet anywhere at anytime using a shared substrate comprising programmable shared network platforms and shared links.

It is now three years since the first GENI proposals were considered by NSF, and it is expected to be at least two more years before GENI is fully funded and several years after that before it is fully deployed and available for use. Bringing virtualization to the Internet is an even longer term prospect, as it requires large-scale changes in business models and relationships among large existing organizations, and such changes rarely come quickly or easily.

This raises the question of whether or not there might be a more direct path to the goal of enabling new networks. We argue that there is, and that the key is to take a fresh look at overlay networks,[1] to see if they can become the key to the creation of new network services and the applications that use them. Overlay networks have already established themselves commercially, most visibly in the context of content-delivery networks, such as Akamai. However, overlays can also be used for a variety of other applications, including VoIP, interactive video and real-time multi-player games. While current overlay networks do not lend themselves well to implementing such applications on a large scale, with acceptable performance, we argue that straightforward improvements in the way overlays are implemented can change that. These changes require no advances in technology and no large-scale changes in business models.

Let's look at what it might take to make overlays a more effective vehicle for implementing new network services on an Internet scale. For concreteness we will focus on PlanetLab [PE02, CH03], since it is the most familiar overlay platform for the networking research community, and because it uses essentially the same basic implementation approach as commercial overlay

---

[1] We use the term ``overlay'' here in the broadest sense, where one protocol is carried over another - and not necessarily over IP.

networks, albeit on a much smaller scale. There are two aspects to be considered - the nodes and the links.

In PlanetLab, the nodes are individual PCs that host multiple applications, each in a separate virtual machine environment. While this has proven to be a useful platform for experimental research, it is a very limited platform for networking or for Internet-scale service delivery. Applications running on a PlanetLab node compete for a very constrained resource pool and have limited ability to reserve resources for their exclusive use. User-space forwarding of packets makes it difficult to process more than 50-100K packets per second (in total over all slices), limiting throughput to less than 50 Mb/s for traffic loads dominated by short packets. Because applications run in user space atop a conventional time-shared operating system, packet processing delays can vary widely and often exceed hundreds of milliseconds, even in systems that are not heavily loaded. However, recent work [TU07] has shown that by restructuring the application to separate the fast path processing from control and exception-handling, and taking advantage of commercially available network processor blades, it's possible to improve throughput by a factor of 50 and to cut delay by more than a factor of 100. Using a scalable configurable platform that links multiple network processor blades and general purpose servers through a 10 GE switch card, it's possible to construct an integrated overlay network node that can handle Internet-scale traffic volumes (>100 Gb/s), while supporting new network protocols and services. Such a system could be used by commercial overlay providers today to deliver services and end-user applications. Smaller scale configurations could usefully be deployed in PlanetLab, greatly improving PlanetLab's ability to host services that carry substantial traffic volumes.

What about the links? To date, PlanetLab has relied on ordinary internet connectivity to link the nodes to each other and to end users. PlanetLab experiments share (statistically) the links with each other, and with other background traffic. The VINI initiative [BA06], which grew out of PlanetLab, is moving towards provisioned links connecting key backbone sites, in order to provide greater control over network bandwidth use and enable provisioning of bandwidth to individual slices. While commercial services that use overlay networks generally reveal little about how their services are implemented, it seems likely that provisioned links play a significant and growing role in commercial overlays as well. MPLS service has now become widely available, with multiple ISPs offering provisioned MPLS links joining customer sites. At least one provider is offering dedicated SONET circuits between customer sites, under the control of large customers [ATT]. In future, it is reasonable to expect customers to be able to create their own provisioned links through some combination of MPLS, SONET and GMPLS. This makes it relatively straightforward to create provisioned links with bandwidths of, say, 2.5 Gb/s and 10 Gb/s joining customer sites, and makes it relatively easy to add new links as needed. The overlay network is free to use the provisioned link as it pleases; for example, it is not constrained to use the IP protocol.

While MPLS, GMPLS and SONET provide an effective solution for backbone links, they do not address access. In the portion of the network between the end user and the nearest overlay node, there is still a role for IP. It's likely that IP will serve in this role and that provisioned links to end users will not be needed. However, it's also possible that the emergence of Metro Ethernet services (using stacked VLAN tags) will enable users of overlay networks to access them through provisioned virtual links, implemented as point-to-point VLANs. Full-featured 10 Gb/s Ethernet switches are now becoming available at costs far lower than conventional routers, making metro-Ethernet inexpensive for carriers to provide and convenient for customers to connect to. Such services can give overlay network providers direct pipes to end users, enabling high quality end-to-end QoS. It also makes it possible to completely eliminate IP, even in the access.

We have argued that the technology to implement far more capable overlay networks exists. There remains the question of how overlay services will be delivered, in practice. The most compelling scenario is through commercial *overlay hosting services*. An overlay hosting service would operate data centers in large metropolitan areas with ample internet bandwidth for user access and provisioned backbone bandwidth joining its sites. It would lease processing blades at its sites to overlay network providers, and supply them with the tools to configure and operate their own overlay networks across this shared infrastructure. To an overlay network provider, there is an obvious advantage to using a hosting service, since it allows the provider to concentrate on the service and the applications it provides, while letting the hosting service deal with the management of physical infrastructure in multiple locations around the world. To the hosting service, having multiple overlay network customers provides diversification and reduces risk.

In a real sense, PlanetLab provides the prototype for an overlay hosting service. A commercial service incorporating the improvements to nodes and links outlined above and scaled up to support internet-scale traffic volumes could provide most, if not all of the benefits of a virtualized Internet. It would allow multiple end-to-end packet delivery services to co-exist alongside one another, providing an effective solution to the problem of Internet ossification. It would make it possible for an individual organization to develop and deploy a network architecture with a unique service model, deploy it on a global scale and scale it up as traffic demand warrants. And the obstacles to deploying an overlay hosting service are far more modest than those that must be overcome to virtualize the internet. A single organization can provide overlay hosting, and while it will require data centers in many locations around the world, it need not own the local network infrastructure. It need only have access to virtual link services that are becoming widely available and can be expected to become ubiquitous in the years ahead.

We close by observing that large-scale overlay hosting services offer a practical long-term solution to the problem of IP ossification. The forces that make it difficult to improve IP today will likely also block future attempts to replace IP directly with a new network architecture. The development of open technology solutions for overlay hosting – that are capable of Internet-scale traffic volumes – can speed the development of such services, allowing more rapid deployment of new network architectures and the applications that use them.

REFERENCES

[AN05] Anderson, Tom, Larry Peterson, Scott Shenker and Jonathan Turner. "Overcoming the Internet Impasse through Virtualization," *Computer Magazine*, 4/05.

[ATT] AT&T press release. "AT&T Unveils Industry's First Service to Allow Customers to Provision Optical Bandwidth on Demand,".`http://www.att.com/gen/press-room?pid=4800&cdvn=news&newsarticleid=23147`, 2006.

[BA06] Bavier, A., N. Feamster, M. Huang, L. Peterson, J. Rexford. "In VINI Veritas: Realistic and Controlled Network Experimentation," *Proceedings of ACM SIGCOMM*, 2006.

[CH03] Chun, B., D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, and M. Bowman. "PlanetLab: An Overlay Testbed for Broad-Coverage Services," *ACM Computer Communications Review*, vol. 33, no. 3, 7/03.

[GE06] Global Environment for Network Innovations. http://www.geni.net/, 2006.

[PE02] Peterson, L., T. Anderson, D. Culler and T. Roscoe. "A Blueprint for Introducing Disruptive Technology into the Internet," *Proceedings of ACM HotNets-I Workshop*, 10/02.

[RH05] Rhea, S., B. Godfrey, B. Karp, J. Kubiatowicz, S. Ratnasamy, S. Shenker, I. Stoica and H. Yu. "OpenDHT: A Public DHT Service and Its Uses," *Proceedings of ACM SIGCOMM*, 9/2005.

[TU07] Turner, Jonathan, Patrick Crowley, John Dehart, Amy Freestone, Brandon Heller, Fred Kuhns, Sailesh Kumar , John Lockwood, Jing Lu, Mike Wilson, Charlie Wiseman and Dave Zar. "Supercharging PlanetLab – a High Performance, Multi-Application, Overlay Network Platform," to appear in *Proceedings of SIGCOMM*, 8/2007