The COOK Report on Internet

© Cook Network Consultants

Changing Role of Peering &Transit in IP Network Interconnection Economics

Development of Bandwidth Cost Management Capability May Be Critical to ISP Survival

Coping with the Internet Core Oligopoly Demands Innovation

A Synthetic Path Analysis Examination of Netflow Data

Volume XI, Nos. 8 - 9, Nov - Dec 2002
ISSN 1071 - 6327

Introduction Highlights

This issue of the *COOK Report* explores IP network interconnection (peering, transit and exchanges) for the first time since about 1999.

While a lot has changed, a lot remains the same. Peering and transit is just as obscure and complex, even to other network professionals, as it ever was. "That is such a black art," said Richard Shockey co-chair of the IETF ENUM working group, when we told him what we were doing. Nevertheless, though a definitive treatise on the state-of-the-art in peering and transit is very likely impossible, we believe that we have thrown considerable light on the subject with the information that we publish in this combined issue.

We have found that the technology, politics, and economics of IP network interconnection differ significantly -- depending on the size of network and geography involved. In North America the state of interconnection is both more evolved and more complex than it is in other global regions. While our contributors to this issue cover the globe, we shall focus in most detail on the structure and evolution in North America.

In North America the Tier 1's oligopoly of peering only with themselves is still well entrenched. Farooq Hussain has written for this issue a remarkably candid summary of the evolution of the Tier 1's peering policy. They are, he says, the Internet Core Networks that announced anonymously on December 5, 2001 their decision to move their peering to Equinix Exchanges. He identifies them as UUNET, Sprint, Cable and Wireless, Genuity, Level 3, Qwest, and AT&T. He also finds their peering requirements to be arbitrary beyond reason. For example, interconnection at OC48 is one thing, but to be forced to do so at 15 locations around the United States is something else again.

Estimates of the capacity utilization of the Tier 1 backbones show them to be lightly utilized at about 15 to 20%. Given this situation Sprint, for example, is undoubtedly quite happy to have SBC buying nine OC-48s. Because the ISP and backbone industry is unregulated, what knowledge we have is sketchy and largely subject to the willingness of folk who both know, and will take the risks of speaking up. Given the state of the industry such folk are few and far between.

Over the past eight weeks, to generate the material for this issue, we have had conversations on a private mail list with more than 30 people who are closely involved in Internet exchanges, peering and transit. Some of these folk have suggested in voice conversation that the oligopoly is engaging in behavior that could blow up in ways that would be very embarrassing for the industry. [For the moment however this is only speculation. See, for example, the sidebar exchange between Miles Fidelman and Sean Donelan on page 78 below.] Whatever happens, peering and transit arrangements have major impacts on the economics of ISP operation. Real money is involved and with real money come power struggles. Certainly, with all seven loosing money and four of the seven (UUNET, Genuity, Qwest and Level 3) either in bankruptcy or in dire financial difficulty, behavior at the Internet Core is not likely to become customer friendly.

On the Inside
Peering and Transit Economics of Interconnection

Contents

Folk associated with the carriers complain that long haul transit backbones are being forced to sell transit essentially below their cost and suggest that something about the system has to change. Of course since almost everything is kept under strict non-disclosure and since these seven players are also either telephone companies, or associated with telephone companies, the picture becomes even more murky due to opportunities for subsidizing operations from the telco side of the house.

In the midst of this uncertainty, the United States government has defined the Internet as a critical telecommunications infrastructure. Yet, one of many ironies of the current situation is that virtually no one in the federal government nor in the financial community knows with accuracy how strained the financial position of each company is. In Europe with the demise of KPN Qwest this spring we saw how quickly a major player can disappear.

But there are mitigating factors to be found in the uncertainty. Thousands of small ISPs survive by dint of hard work and because of what we contend are, for them, favorable economies of scale. They are defining analytical tools that we discuss in detail in this issue. They are weaving their positions at the edge of the network into a mesh of cooperation that is likely to provide resiliency for the Internet as a whole as shockwaves from the collapse at the center propagate.

Consequently, the lack of detailed knowledge about the economic condition players at the center may not be as problematic as we had believed. Discussion with the 25 contributors to this issue (listed on page 6) has shown us that by the time the US and other governments collected the kinds of data that the telcos are required to submit data on ISP interconnection and transit economics would be out of date.

Ascendancy of Current Tier 1s Being Challenged

The tectonic plates of network traffic and power are shifting with the economic uncertainty brought on by the industry crash and the increase of cable modem and DSL traffic. Given the extraordinarily low cost of bandwidth and the existing investments of US carriers in some of the fiber players, we can expect very soon to see a build out on the part of these carriers into peering at Asian and European exchanges. In this sense a lot of effort will be put by large players into moves to enable them to avoid paying transit fees to the currently seven largest global backbones (Tier 1). In doing so, the likely outcome is that these new comers will eventually either replace or join the Tier 1 oligopoly.

They are, in effect, climbing a peering "ladder" where as their bandwidth increases and they peer with each other and can get peering with larger players, they are likely to depeer with smaller players whom they feel they now no longer need and believe they can sell transit to. Thus although the plates of peering are shifting, the fundamental premise is likely to remain one where players peer only if their aggregate traffic is approximately equal. However peering is motivated as much by politics as economics. Therefore, it would be a mistake to think that these premises apply with equal force the world over.

Bill Woodcock has shown that a case can be made for the position that it makes sense for a larger network to accept traffic from a smaller network that terminates on the larger network. However the larger players are still firmly of the opinion that size differences in network traffic are there to be exploited by the larger and presumably more powerful network. Not surprisingly this view is firmly rejected by the smaller players. Therefore as the new broadband based networks move to extend their peering infrastructure around the US and across oceans, they are likely to act increasingly like the Tier 1 oligopoly they want to replace and seek to sell transit to rather than peer with others who haven't grown as fast.

What is unknown is how good a job how many smaller players can do of extending peering with other small players through use of the approaches and methodology outlined by Woodcock in the long interview in this issue. At an abstract level Woodcock's views that peering is a good thing seem to be understood and accepted by the large players as well. The problem for them seems to be one of "good for whom?" along with the belief that, once you get to a certain size (and we might add business model), they don't scale.

Therefore as the new broadband based networks move to extend their peering infrastructure around the US and across oceans, they are likely to act increasingly like the Tier 1 oligopoly they want to replace and seek to sell transit to rather than peer with others who haven't grown as fast.

Although the make up of the Tier 1 backbones may change, the existence of a handful of global networks at the top of global IP transit food chain is not likely to change in a serious way. Perhaps the most significant unknowns are how large and sustainable the Tier 2 doughnut around the Tier 1 transit providers can become. Also it is uncertain whether the benefits derived from joining the Tier 1 club will be great enough for the new entrants to find that the results actually pay back the efforts and expense in infrastructure building necessary to accomplish their goals.

Ren Nowlin reminded us that peering terms are not set in granite. "Future forward crystal balls are vague. Never forget depeering is a variable in business plans. Some peers are migrated for free and others for a fee when networks evolve and move out of IXs depending on perceived value at that stage by the two parties. SBC does not plan to depeer today but Level 3 didn't either at this stage of network development. Peering agreements, and terms, are not static."

Players like SBC and Level 3 have to contend with billions of dollars of debt. Consequently it seems very likely that whenever management believes that depeering would be likely to produce additional income, depeering will be likely to occur.

It's the Economics of Networks, Stupid!

With the industry collapse, the tightly run economic aspects of one's business matter in a way that they did not during the bubble when money flowed freely. In this context, some important new tools and concepts are being developed in the peering and transit arena. In addition to Farooq's examination of the current global climate at the Tier 1 level, this issue looks at the development of some tools and methodologies to manage, much more cost effectively, an ISP's interconnection costs.

Since about 1998 Cisco (and now Juniper) routers have had the capability of giving the users what is called Netflow data. Use of the Netflow data can give significant information about where a network's traffic is going, including what autonomous systems the ISP's traffic flows through to reach its customers. Various efforts are underway. A small handful of ISPs of differing sizes are beginning to use tools by companies like Route Science and SockEye to do load balancing of their up stream connections in real time. Some folk are also beginning to build tools using Netflow data to help them most cost effectively analyze how to do their transit and peering in the first place. One such effort is by Jeffrey Pappen, Peering Coordinator for Yahoo. It is called TUNDRA, The Ultimate Netflow Data Realtime Analysis, and was presented by Pappen at NANOG in October 2001. See www.nanog.org/mtg-0110/ppt/tundra.ppt Another is by Martin van den Nieuwelaar and is called Network Intelligence. See http://www.networkintelligence.biz Stephen Stuart at PAIX has looked at Netflow approaches and applied some of them there.

Another, and likely the most important, effort has been developed by Bill Woodcock of Zocalo.net, and Packet Clearing House and Alex Tudor at Agilent technologies. This is discussed in great detail in a 16,000 word interview with Bill Woodcock on pages 12 to 27of this issue. We note that Woodcock's approach is the subject of a patent application. In the language of the legal department of Agilent. "Some of technology described in this article [our interviews] has been claimed in one or more pending patent applications that are owned jointly by Zocalo and Agilent Technologies, Inc., each of whom is free to license."

On September 2 Bill explained the patent issue more clearly: "The deal between me and Agilent was that I would do a "technology transfer" and tell them what code to write. They wrote the code at my instruction, and we jointly filed a patent application to preclude someone else's making proprietary claims against it. The content of the patent disclosure and application are open source, and free for public use. Again, the purpose of the patent application, is to protect open-source use of the algorithms against future claims that they're someone's proprietary invention."

Finally we note that from, an economic point of view, the most important component of what Woodcock is doing is what he calls Synthetic Path Analysis. That analysis is publicly discussed below for the first time. It has not yet been tested in production use over a period of six months to a year's time. This should be done. Until it is, we can say only that from an intuitive point of view it makes strong sense. Certainly ISPs should be looking at experimenting with it.

We think that what Woodcock says is extremely significant. Among other things, he points out that the Tier Ones, by peering in their tight oligopoly, may have rendered themselves irrelevant. Why? Because the smaller networks with rich peering are beginning to build such a well connected donut around the Tier 1's that one may be on the verge of being able to deliver one's traffic to all of the destinations to which it needs to go without relying on the Tier 1's for transit.

In short, ideas and methodology are evolving. The new topology looks more like a multifaceted geodesic dome than like a dozen global backbones with networks hung off them tree-branching fashion. Critical components of this new topology, in addition to peering and transit circuits, are the hubs into which the circuits are attached. These are the more than 300 exchange points around the world that facilitate network interconnection. There it seems likely that newly cost-conscious ISPs will increasingly deploy Netflow data methodology to model their traffic and decide from the results where and with whom to interconnect. When we described to the technology director of a large CLEC what they do, he said: "that sounds pretty much like how we model and plan our long distance circuit interconnections with the PSTN. Do it right and you are profitable. Do it wrong and you are history."

Of course what is right and what is wrong is likely to be a function of where one is in the industry. Below the Internet Core oligopoly one has a group of very large players who are themselves under stress and therefore must find strategies for survival. This group includes AOL, SBC, Yahoo, Shaw Cable, France Telecom, Equant, Verizon, Bell South and many foreign carriers like Telstra and NTT. See for example http://www.sbcbackbone.net/peering/ They have their own ideas about peering and given cooperation by them we may be able to begin to explore them in the future. (No promises because such cooperation is not yet firmly in place.) Beneath this group are the several thousand smaller ISPs with which Woodcock works. We believe that both these groups could profit from Woodcock's understanding.

Future Direction

Vertically integrated local phone companies are forced to have a business model that squeezes every drop of money from every drop of traffic if they are to get enough income to pay the interest on their bonds. Forced into this procrustean bed, they will find that the economics of an industry concentrated on broadband with user control at the edges and less and less concern with extracting rent from the content are capable of benefiting from the declining costs of bandwidth and the hardware necessary to support it. The local companies will either adapt to the new model through bankruptcy or find through political interven-

tion a way to bail out their debt ridden foolishness and go merrily on as before promoting bandwidth scarcity and ensuring that the telecom industry in the US in the 21st century shares the fate of the steel industry in the 20th century.

The questions facing us are no longer just ones of technology. They are ones of economics and policy. How we answer them will determine whether our information technology industry prospers or stagnates. One thing that we find fascinating about Woodcock's peering and transit methodology is that it may begin to provide an answer to the question of how asset based telecom and the fringes of the network can stitch themselves together to begin to replace the long haul carrier networks.

We begin this issue with Farooq Hussain's overview of global peering and transit. There follows the two part interview with Bill Woodcock. Bill suggests a very interesting world view where ISPs that are clueful and careful can make themselves competitive by avoiding interconnection with their ILEC and backhaul circuits to appropriate exchange points. They should peer as much as possible at an exchange where the cost of interconnection is as cheap as possible. Reliability is not critical because, should peering sessions fail, traffic can be delivered through the transit providers. Since transit is critical one should connect to at least two transit providers ideally at different exchanges. One figures out in each specific instance how to do this using the path analysis tools and then one simply does it.

Certainly, for the multiply-homed ISP, this is a viable model. However, when one gets to an entity the size of Verizon, SBC, Bell South or one of the MSOs with tens of thousands of cable modem customers one hears a different tune. There it is a huge difference in scale including national operation and perhaps a gigabit or even multiple gigabits per second in bandwidth requirements that prevents reliance on the ability to cutover quickly, when a problem arises, a large amount of peering bandwidth to a transit provider. The amount is such it is said that no transit provider could be counted on to accept instantaneous cut over without itself being swamped.

Yet these big up and coming want-to-be Tier 1 networks do peer. In fact most peer very assiduously. The ILECs in general still have only one transit provider. SBC uses Sprint. Verizon relies on Genuity and BellSouth on UUNET. We are told this is because of FCC inter lata long distance restrictions. As more and more are granted long distance authorization in their respective states, we are told this will change and the LECs will very likely begin building their own backbones. Given SBC's investment in Williams, BellSouth's in Level 3 and Verizon's in Qwest, building backbone infrastructure in the US and extending lightwaves to Europe and Asia for peering purposes would be trivial.

We don't claim to have special insight into the LEC or MSO operators peering mindset. Even were the LECs of the persuasion to do as much peering and little transit as possible, we suspect that such an attitude wouldn't stand up very well in the face of the LECs need to maintaining revenue and pay down debt. The same would go for their cable brethren. Despite the fact that transit bandwidth is by all standards exceedingly cheap having fallen in cost by anywhere from 80 to 90% in about two years, the LECs appear to be ready to make capital expenditures that go beyond transit costs to buy infrastructure. They say it is to have greater reliability and flexibility within their own networks. Many smaller independent ISPs feel the real motivation is to bury them.

The overarching problem looks to be that the Stupid Network doesn't yet seem to have a sustainable business model. If costs are indeed focused mainly at the edges what we have seen in the willingness to invest in user owned assets at the edge is reassuring. The problem is that there seems to be an unfundable unsustainable vacuum in the middle that no one has figured out how to deal with. But figure it out we better for as Roxane Googin points out until we solve "The Paradox of the Perfect Network," the telecom and IT industries will remain flat on their back. The result will be a permanent recession and an end to growth.

Editor's Note: Reader's will observe that our Executive Summary (appearing for this issue on pp. 117-118) is quite short. This is because material in this Introduction also serves as a summary of global snapshot of the state of Internet interconnection we have just compiled.

A Footnote on Comparative Size of Internet and Telecom Industries

On Sept 27 Andrew Odlyzko: Gordon's mention of subsidies from the telco side of the house brings up something that Bill Woodcock and I have mentioned before on this list. The Internet is still far too small to pay its way. In the US, the whole Internet service business brings in revenues of around $15 billion per year (and only a small part of that comes from the wholesale backbone transport), while the whole phone industry brings in around $300 billion. The low prices that we are seeing are to a large extent the result of the irrational exuberance of the financial markets, which were assuming that data revenues were going to explode. (Jack Grubman, the famous, or, to be more precise, currently infamous, Wall Street analyst/(cheer leader)/huckster/..., is defending himself against accusations he misled the investing world by saying he honestly believed telecom spending as a fraction of GDPwas going to double in the next few years, with all the growth coming from data services.) That simply has not happened. Even if we throw in dial access revenues, we discover that the Internet is something like a $30 billion dollar a year business in the US. Cell phones alone are at something like $80 billion a year. Eventually we will surely move to an environment where broadband data transport dominates revenues, but that is still far into the future, and in the meantime we have this awkward transition to manage.

Contributors to Peering & Transit Symposium

Dileep Argawaal, Founder and CEO of Worldlink, Kathmandu Nepal

John Brown, CEO Chargres Technologies

David Diaz, Peering Advisor to BellSouth.Net, former CTO NetRail

Ralph Doncaster, founder I-Stop.com Network

Pedro Ferreira, Doctoral Student (transit & peering), Carnegie Mellon

Avi Freedman, founder of Netaxs and Chief Network Scientist at Akamai

Dan Golding, Effective Sept 27, Peering Manager AOL.

Roxane Googin, Telecom and IT Equities Analyst

Mike Hughes, Network Architect, LINX

Farooq Hussain, Partner at Network Conceptions and former Sprint NAP PI

Joe Klein, Peering Coordinator Adelphia

Kurtis Lindqvist, CEO Netnod, former network Architect Peering Coordinato, KPNQwest

Francois Menard, Project Manager, IMS Conseils

George McLaughlin, Director AARNet, Australia's Academic and Research Network

Keith Mitchell, founder and CTO Xchange Point

Lauren Nowlin, Peering Coordinator SBC

Andrew Odlyzko, Director, Digital Technology Center, Univeristy of Minnesota

Jere Retzer, Senior Manager, Next Generation Networks, OHSU & Co-founder Northwest Access Exchange

Philip Smith, Consulting Engineer, Office of the CTO, Cisco Systems

James Spenceley, Network Architect, COMindico. Australia

Stephen Stuart, VP Engineering, PAIX

Martin van Nieuwelaar, Network Intelligence Software

Wouter van Hulten, Principal, Interxion.com (European exchange point operator)

Alexander Tudor, Researcher Agilent Laboratories and Collaborator with Bill Woodcock

Phil Weller, CTO of FastNet

Bill Woodcock, Research Director of Packet Clearing House and founder of Zocalo Networks

Why Peering and Transit Is Badly Broken at the Global Tier One Level

How the Process Evolved into the Oligopoly of the I-Core

by Farooq Hussain - Highlights

Editor's Note: We have worked carefully with Farooq Hussain through about three drafts of this overview article and much appreciate his patient response to our questions and editing. The result, we believe, is a very important piece. We can now place our interview with Bill Woodcock in a global context that enables a much better understanding of what Bill is doing.

When we asked Farooq to introduce himself, he said: I started out with the NSF International Connections Manager program at Sprint and with SprintLink the company's commercial IP backbone. I was the Principal Investigator for the Sprint NAP and moved shortly after the NSFNET transition from Sprint to MCI joining the team directed by Vint Cerf. I've been involved with peering policy since the planning for the NSFNET transition both with Sprint and MCI. I left MCI just prior to the completion of the merger with WorldCom having worked on both the merger plan with BT and subsequently WorldCom for the Internet components including a brief time with Concert when MCI and BT were planning an integrated global IP backbone. I was with AGIS for a little over a year helping to establish a business relationship with Telia of Sweden who subsequently bought AGIS out of bankruptcy. Currently, I'm a partner in a research and consulting firm Network Conceptions together with Phil Jacobson [also an ex-MCIer]. We're focused on providing independent research to institutions, vendors and carriers internationally. As an advisor to carriers I'm actively involved with issues impacting peering policy internationally.

Farooq Continues:

The focus of the much discussions that will be published by Gordon in Part Two (December issue) has appropriately been on the techniques that might be used by ISPs to assess the best peering relationships for them. Since there are costs associated with peering these need to be compared with transit costs when business decisions are to be made. It is interesting to note that only a small number of ISPs are thought to be using Netflow and AS path information in the ways outlined in Gordon's interviews with Bill Woodcock. Thus the opportunity to consider the techniques and the ways in which ISPs might find it easier to obtain the tools and to apply them has I believe, been a very valuable contribution to this discussion.

ISPs have always had to evaluate how to gain efficiencies in performance and cost. But, in the end, I think that even when many more of them are better able to determine peering optimization, their business model will impacted not only by the cost of delivering packets but by broader issues influencing their business model including what they can charge for services. Of course, these influences are very different depending on region and country and the level of competition in the market.

For a developing country that has been opened to competition, the benefits of peering with competitors include keeping costs of international transit down. But the proportion of in-country traffic is as a rule a small fraction of international volume for such countries. It is also not unusual to find that ISPs that have purchased international transit to be inclined to leverage competitive advantage over those who have not. The plain fact that peering cedes some perceived competitive advantage is hard to overcome. Usually the have-nots have to rally together against the haves and peering at present is certainly one good way to do so. If, as your interviews that follow below with Bill Woodcock show, there are techniques that can be applied to determine the benefits of peering between ISPs, so much the better.

How We Got to Where We Are

We have what seems to me to be a thoroughly broken situation. I would like to look back and trace how, the dominance of the seven networks that I refer to below as the Internet Core (or I-Core for short) arose. Three early factors determined much of the I-Core's rise to dominance after 1995. First there was the central role played by the NSFnet backbone in the development of the commercial Internet between 1990 and April 1995. Second there was the international connections manager program run by Steve Goldstein at NSF during this same period. Third was the fact that tariffs for leased data connections were generally cheaper to run from Europe or Asia to the USAthan from one nation in the European or Asian region to the others in the same region. These factors meant that it was economically feasible for the NSF to share with countries in Europe and Asia the cost of running a high-speed link to the US.

Before April 1995 when the NSFnet backbone was turned off, several dozen foreign research and education were connected to the NSFnet backbone. Traffic from one foreign R&E network to another was deliverable only by flowing across the NSFnet backbone in the US. This had the significant result of making the United Sates the hub of the early global Internet. When the NSFnet backbone was decommissioned in April 1995, backbones operated by BBN, UUNET, MCI and Sprint took its place. Domestic traffic from within the US came into NSF mandated NAPs or network access exchanges where it was sent settlement free onto these early commer-

cial backbones. Traffic from Europe, Asia and the rest of the world also flowed into these access points and back out. Much traffic that was destined to stay within Europe or within Asia could only reach its new destination within the region by flowing into the US and back out. The four commercial backbones in the US were critical components of a global hub. They acquired early on large flows of traffic that they were responsible for inter connecting (transiting). With this traffic they acquired the critical marketplace mass of what would come to be referred to as Tier One peering and transformed it into what Bill Woodcock calls the oligopoly of the 'doughnut hole'.

To summarize: back during 1995 and 1996 this Inter-regional traffic (inter European or inter Asian traffic) was once considered to form a significant component all traffic exchanged between countries from the rest of the world and the United States. That is to say, of all the traffic flowing through the US backbones, a large part of it traveled to the US and back only to get from one part of Europe to another or one part of Asia to another. However, this US dominance in the role of European and Asian traffic exchange did not last long.

When networks from other countries connected to the United States, they were concerned about the inefficiencies of traffic exchange. For instance, in the European example traffic from the UK to Germany would have to be to be exchanged in the US. The first issue was that the purchase of capacity [an international circuit] between Germany and the UK was more expensive and harder to come by than it would been to purchase capacity to the US. In the case of Europe, once new fiber networks and competitive carriers began to deploy infrastructure between 1996 and 1998, inter-European traffic no longer needed to be exchanged in the US. Thus beginning a few years back (approximately), European networks, now well interconnected on the continent, have been trying to establish peering relationships with the so-called Internet Core [I- Core] group [WorldCom/UUNET, Sprint, ATT, Level 3, Qwest, Cable & Wireless, Genuity] and others to terminate US bound traffic from Europe in return for carrying European bound US traffic.

[**Editor's Note**: The Internet Core reference is to the Equinix Internet Core Exchange service announced in a December 5, 2001 article in TelephonyOnline.com. http://www.industryclick.com/Microsites/Newsarticle.asp?newsarticleid=262327&srid=10164&instanceid=5700&pageid=847&magazineid=7&siteid=3%20]

Ignoring the Europeans

In discussions with their European counterparts, the large backbone US ISPs have generally maintained a position that the balance of in-bound [from Europe to the US] versus out-bound [from the US to Europe] does not justify a peering relationship. European networks generally agree that this gap in traffic is closing particularly as content value in Europe grows. But there still exists a traffic disparity. Whether the traffic imbalance justifies a denial of peering in the US is open to question especially as the US networks, naturally enough, find it beneficial to maintain peering relationships in Europe with networks to whom they deny peering in the US. In South America and in the Asia Pacific region these imbalances are even more pronounced.

For US networks the issue is simply one of protecting a legacy source of revenue. Before in 1997 US regulation permitted international carriers to wholly own capacity to the US, these carriers had to pay their US network service provider for the international circuit for the US half and meet the costs for their half themselves. International carriers wanted the US networks to meet the costs of the US half which also didn't make any sense since any US network would, in theory, need connections to all other countries while any other country could get by with a single connection to the US. The issue was very fractious between US and international carriers and lead by 1997 to government and institutionally sponsored initiatives to derive settlement-based models for the exchange of Internet traffic which I discuss further below.

Despite these European changes, patterns in the rest of the world have not changed much over time. Despite initiatives to establish regional network infrastructure in the Americas for example, the main flow of traffic from countries there is to the US and inter-regional traffic volumes don't justify the building of networks to serve regional needs. Similarly, in Asia efforts to send European bound traffic directly there instead of via the US, though established, are not competitive with trans-pacific routes. Also many countries have found that the inter-regional traffic within Asia doesn't justify connections between countries in the region. However some have certainly been successfully established. For example, those between Japan and Korea. Hong Kong's importance as a regional hub is in a significant part due to access to trans-pacific routes.

In contrast in Europe inter-regional traffic has benefited from extensive development of inter-country routes made easier from having well-populated countries close on the same landmass - from Moscow to the English Channel. Equally, it is no longer true that the costs of connecting Amsterdam to London would be greater than Amsterdam to New York. It's hard to see how the Americas or Africa can deploy routes with the equivalent competition since they also do not operate as an economic group in the manner of the European Community.

The Central Question and the Oligopoly of Seven

All this said, the central question for peering policy, taken on a global scale, as opposed to policy for the domestic USA, still revolves around who has to pay to get their traffic to the United States network infrastructure and who does not? And related to this one should ask what is happening within the US environment that as a consequence of recent bankruptcy and consolidation in the telecom sector might significantly impact the development of IP networks both within and outside the US?

In my perception as things are, seven networks [WorldCom/UUNET, Sprint, ATT, Level 3, Qwest, Cable & Wireless, Genuity] currently operate an oligarchy dominating peering in a manner detrimental to competition. A significant number of this group are in difficulties as businesses and may not survive intact for very much longer. Behind them are companies such as AOL and SBC who though close to being full members of this group are not quite there - yet. The impact of a peering oligarchy is a significant impediment to competition both for the domestic US market and internationally whether or not the size and composition of the group is marginally impacted by bankruptcy, consolidation or new members.

The structural framework for peering with these networks has its origins in the requirements set out for US national level IPbackbone networks at the time of the NSFNET transition which I would argue are no longer appropriate especially as they have been used in the past couple of years to create a monument to restrictive practices - effectively eliminating all but the I-Core group from peering with one another. Equally, these I-Core networks have worked around peering and transit with "paid peering" relationships for certain networks while never formally acknowledging that this relationship exists. [**Editor's Note**: Paid peering is the ability to send all traffic from the network purchasing the peering that is terminating in the network from which peering is purchased and the agreement says that no transit is to be provided by the paid peer to the network purchasing peering.]

I think that its been a problem from the beginning for US networks to try to explain to international networks that they are not able to charge customers in the US for international access if you like, whereas the networks of other countries felt that it was unreasonable for them to bear the entire costs of connecting to the United States and pass this on to their customers. But the domestic and international business directions of US networks have resulted in extraordinarily restrictive peering policies with I guess Sprint and Cable & Wireless in a photo finish for first place as the most absurdly restrictive from the perspective of how shall I say this? - 'informed opinion'.

First, let's look at the domestic situation in the US. Networks here began to be of value to larger players solely because of their peering agreements. For example, as Level 3 failed to make any progress with negotiating peering agreements with MCI WorldCom and others, it found that it could get the peering it needed only by buying a network that already had it. In this case, its acquisition of GeoNet conveyed much needed peering to Level 3. There has been a striking some might consider it alarming, consolidation of peering relationships as a result of the acquisitions, bankruptcies etc. in the past two years.

With a much bigger gap between the I-Core networks and others in the US market, the I-Core group, as a whole, progressed towards more and more restrictive requirements for peering. The hurdle for even a large CLEC for example to meet the peering requirements of a typical I-Core member is just totally unrealistic from a commercial perspective. These requirements have included substantive geographic footprint, multiple very high capacity redundant interconnections and more. Almost all of the I-Core networks have also all withdrawn from public peering exchanges and certainly none of them is willing to establish a new peering relationship at any public exchange so far as I've been made aware. In December 2001 Equinix announced that a group of Core networks would be coming to the Equinix Exchanges. Sources say that they will come to one or two other exchanges as well but that they will continue to interconnect only among themselves. Coming to the exchanges in no way implies a decision change the oligopoly's mode of operation.

The I-Core group and close followers have essentially abandoned the concept of meeting peers at exchange points. For example, I've heard from several sources that SBC is in the process of shutting down the PacBell NAP and is similarly downplaying the AADS NAP in Chicago. Sprint killed their NAP to new customers a while back and WorldCom disingenuously made an announcement a couple of months back about expanding the MAEs - but of course you can't peer with UUNETat any of them. So we have a default direct interconnect requirement for all the I-Core group. Though this is actually an attractive requirement for larger international networks, should they want to attempt it a whole fabric of requirements unfolds for which the only rationale seems to be an exclusionary one. For example, the large networks with lots of traffic, were they able to interconnect, would have no problem in doing so. But the only problem is that they are then told that their direct interconnection would have to take place in as many as 15 different locations in the US.

On the face of it you have in some cases published peering policies. These policies place extraordinary business costs on any domestic or international network. They do so to the point that attempts at compliance make no commercial sense. Furthermore, these are mixed up with arbitrary arrangements for 'paid peering' or special deals on transit which I'm led to believe many of the I-core group maintain.

Basically, the I-Core networks have established a range of options to keep their transit and paid peering revenue flowing even though, for the international business, the majority of the revenue associated with very high value markets such as Japan, Korea and a few other Asian countries and Europe have disappeared as those carriers have purchased their own capacity to the US. The most seriously impacted networks in the current situation are those of the developing countries but it's also rather serious for the larger international carriers also who are pretty much captive to the framing of domestic peering policy in the US resulting in bills to the tune of hundreds of thousands of dollars per month in some individual cases.

Some of the I-Core networks also segregate peering relationships with networks

by international geography so that a network that has a peering relationship in Europe for example finds itself having to pay for transit or paid peering in the US. For larger networks transit and paid peering costs run into very large numbers particularly, when taken as an aggregate, they must amount to many tens if not hundreds of millions of dollars a year. In my view this is really an unstable situation. On the one hand we're learning that capacity swapping and related revenue booking has been identified as having been conducted fraudulently by a number of carriers. It's wholly my opinion, but at some point forcing networks to pay revenues by racking up extremely restrictive interconnection requirements is the other shoe just waiting to drop.

Peering between IPnetworks was founded on the idea that the costs of equipment and facilities would be in approximate parity between partners and that the cost of carrying each other's traffic was also a wash where the volume of traffic was near equal. Additionally, parties agreed that there would be no third party transit and no snooping. Some, though by no means all of the seven networks mentioned above publish their peering requirements and all of these reflect a condition that the network requesting peering meet criteria that essentially require it to be of comparable size in network and investment. This is not fundamentally different from the beginning premise but now at a level to exclude just about everyone but themselves.

Why Existing Policy Is Broken

My concern is that even with implementations such as the one proposed by Bill Woodcock in his *COOK Report* interviews, all networks, whether they are an ISP in Madagascar, an international carrier or service provider in the US, will find their business model being significantly impacted by the now outdated policies as represented by the I-Core group of networks mentioned above. In addition to the proposal to provide techniques and tools to assess the best peering relationships and to be able to evaluate peering vs transit costs, it is also important to recognize that existing policy is broken. An overhaul of policy that recognizes the technical rationale for networks of unequal size to exchange traffic and which places reasonable mechanisms for establishing the monies to be exchanged between the interconnecting networks is also sorely overdue.

Over the years I've been disappointed with how unsuccessful we've been as a community in working to resolve the fundamental problems with peering that I've outlined above. From the early 1990s under the leadership and lobbying of Telstra of Australia, institutions such as APEC and the ITU had tried to encourage the rest of the world to adopt a settlement based model for the exchange of Internet traffic. The program created to do this is called International Charging Arrangements for Internet Services (ICAIS).

Since the initial zeal of Telstra peaked in 1997-98, the torch has passed first to Singapore and most recently to China. Ironically, since about 1999 Telstra has abandoned its former position and no longer supports the pursuit of a settlement based regime. Nowadays the main argument for a settlement-based model has become a development issue where revenue from settlement for the exchange of Internet traffic passes to developing countries to help build their infrastructure. From my view I believe that the development issue is important but that economic development funding rather than a settlement model should be used to support it. Especially, as after a decade of sporadic attempts, no generally acceptable model has been identified for settlement-based exchange of Internet traffic.

Even though US networks have acquired ISPs in other countries and claimed to have established a global presence, they've still not demonstrated a recognition of the general benefit to the global marketplace and to themselves of using their global presence to overhaul their business model for peering. I believe that the industry sector now has the chance to address the issue within the framework of the broad re-structuring and consolidation that is taking place with or without government and institutional pressure to do so. For a small network just starting out, the decision between peering and transit is always going to favor transit as long as these services are available in a competitive market rather than one dominated by a few players whose peering policies unreasonably prevent new competition.

In many respects the complexities of the various peering arrangements that have developed have occurred because the cost of purchasing a transit connection from the seven members of the I-Core, and a few others, had become commercially unsustainable. Faced with this problem, we also had the development of Ventures such as InterNAP. InterNap established services based on proprietary technology to offer 'best route' based on transit services purchased by InterNAP from at least three of the largest IPbackbones.

The problems for InterNAP, as I understand them, came primarily from the fall in the price of bandwidth after they'd negotiated long term agreements from suppliers which they were not able to re-negotiate. However, at present most operators appear to managing a mix of transit, paid peering and peering relationships. In the mid 90s the incentive for US networks other than the three original backbones UUNET, MCI, and Sprint to seek zero settlement peering agreements, was driven as much by investor requirements that considered doing so essential to their ability to bring out an IPO as by that the cost of transit was always falling.

But for larger networks the costs of transit are a very significant component impacting their business efficiency. Some networks such as AOL for example might be able to leverage their market power to gain peering relationships or to otherwise drive down the costs of transit [If you don't give me peering, I take my dial access business elsewhere] but this type of leverage is not available to the vast majority of ISPs.

Bill Woodcock Explains New Framework on Which to Build Peering and Transit

Highlights Editor's Note: Bill Woodcock is research director of Packet Clearing House, a non-profit research institute dedicated to understanding and supporting Internet traffic exchange. Bill has operated regional and national Internet networks since 1989, and currently spends most of his time building Internet exchanges in developing countries.

COOK Report: Where are trends in the ISP market headed? How does one determine what is a finically viable ISP at this point in time?

Woodcock: I think that one thing we have been seeing is an evolution away from regionals. There aren't a lot of ISPs that cover areas of from half a state up to five or six states. Whereas if you look back to the end of the 80s when things got started, all the networks were effectively regionals.

It became clear very early on that a regional size was not effective for dial up ISPs. You needed to have either one pop and no backbone costs or you needed to be nation wide. Either way you could make a living. If you tried to wedge into some middle territory you would either have people below you with lower costs or people above you with greater functionality.

Evolution of ISP Business Models

I am less clear about the future for leased line ISPs which I regard as more cloudy. I have always felt that I understood the business that I am running but I am much less clear about my insight into the motivations of other people. The vast majority seem to have been running at a loss.

COOK Report: Would you elaborate more? I assume the ISPs that you and Avi Freedman were running for example had been profitable.

Woodcock: I think that Avi and I generally managed to keep things profitable but I know that a I had quite a few customers that were themselves leased line ISPs and were loosing money over the long term. I know that many of my competitors were losing money and eventually succumbed to it. In the case of tier ones, the big phone companies, we know that they were subsidizing their losses from other parts of the business.

COOK Report: I believe that in the case of Netaxs, from 1992 through at least 1997 or 98, Avi financed his expansion from cash flow without having to borrow. But does there come a point where you can no longer do that?

Woodcock: I know that this was certainly true for me as well. The problem for me was that the market was becoming less and less qualitatively differentiated. The people who were doing a really bad job of things were able to satisfy most of their customers most of the time by simply throwing more fiber at things. They would go out and get more capital to get more fiber and by having more bandwidth to throw at their operations they got by.

Now if you look at the last couple of years (2000 onward), there has been less and less of an end user perceivable qualitative difference between good ISPs and bad ISPs. Being good at what you do has become less of a saving grace. Better engineering and a better business plans, as more and more money was thrown at the market sector, didn't benefit those with "clue" in the manner that one would have hoped. Because of the way the market was skewed during the bubble, an ISP with a bad business plan and bad engineering could still have been popular with customers by having a lot of bandwidth at a low price.

On the one hand it was gratifying to see the whole Internet business boom. On the other hand it was a little heart rending to see people who knew the Internet and did high quality engineering get run out of business by people who didn't know the Internet and couldn't care less about doing good engineering but who had large quantities of money at their disposal.

What I don't know is whether we will have a return to people who really care about good engineering or whether end users' expectations have been permanently lowered and they are multi homing as a result. In any case, phone companies having too little clue and charging too little money, screwed up more often and lost the trust of their customers who then started to multi-home. There was, as a result, a kind of downward spiral.

The Multi-Homing Downward Spiral

In the midst of this you could differentiate by saying: I have enough clue to get my customers up and running using BGP and multi-home them to someone else. Zocalo did a lot of this in the last few years. At the point where I turned it over to Joe some 40% of our customers were multi-homed. We never lost a multi-homed customer. They became very stable customers at that point.

The problem was what was driving all this multi-homing was a fundamental distrust of ISPs. The downward spiral went: teleco screws up; customer says telco isn't reliable any more. What do we need to do about this? We need to get service from someone else. But we don't want to spend anymore than before. Therefore we need to pay half as much to each of two providers. However, even if the customer only shoved half as much through each of his multi-homed lines, for the providers having to maintain them, the overhead was still there. Of course the customers' use of bandwidth was rapidly growing and therefore the amount of bandwidth sent down each new line did not decrease for long.

When customers started doing this the result was that margins got very very tight. With margins getting tighter the ISPs had less to spend and this in itself contributed to the downward spiral. The ISP might then put less work into engineering and quality assurance. As a result it would screw up again. The spiral would then get reinforced as the customer might set out to get even more redundancy in its net connections.

The problems in the supply chain however, were sometimes such that when the phone company screwed up, the ISP would take the blame and as a mid range ISP remember that you dare not single home to just one backbone. It can be argued that the regionals provided the Internet's reliability and as the regionals went away you get more and more of a situation where huge backbone carriers were competing against small local providers and neither one of them very well multiply homed. The customers then wound up having to do the multihoming. This demands a lot of technical ability all the way down from the Tier 1 to the customer where in the real world it simply is not generally found.

COOK Report: Perhaps about 1998 Dave van Allen was claiming that Fastnet provided reliability by homing to five or more of the Tier ones. It sounds like you are saying that this model never really took off.

Woodcock: InterNap was really the final incarnation of this model and made the final big push in this direction. But InterNap's failing turned out to be that they didn't do any peering of their own. Therefore they had no way of off loading traffic without paying transit prices.

InterNap essentially was an ISP that had no backbone and no peering. They had POPs which were multi-homed. If you bought connectivity from them, you were buying transit from one of their POPs. At their POPs they would have connections to a bunch of different backbones. A lot of people did things kind of like this. We had POPs that had rather minimal connectivity between each other but a lot of transit going outward. There is no reason why it doesn't work as long as you don't get big enough on the radar screens of the Tier ones for them to realize that you are doing this and figure out that they have you over a barrel.

COOK Report: What about Abovenet?

Woodcock: Abovenet started a bit earlier. It was very aggressive about getting peering. It explains why their growth curve was so fast.

COOK Report: Weren't they an aggregator of smaller ISP bandwidth? By peering with everyone didn't they gather together bandwidth that could be parlayed into peering with the bigger players?

Woodcock: There is nothing inaccurate about such a statement. However, I would not say that that was the real significance in what they did. Rather, at a time when Tier 1s were shutting down public peering, Abovenet stepped into the gap that such idiocy created and took advantage of it. They stepped in and said: these guys are cutting their own throats. We on the other hand will peer everywhere with everyone. Doing this will give us more to sell. This was a the period when MCI had like a five six or seven month sales moratorium because they had no more bandwidth to sell.

The Sum of All its Peering

All that any ISP has to sell is the sum of all its peering. This is as true for the Tier Ones as it is for everyone else. We are talking about the sum of the outbound bandwidth. So for the Tier 1s it is only peering, and for the smaller guys it is the sum of the peering and transit. That being the case, someone who has restrictive peering requirements is simply not going to grow as fast as one who aggressively peers. They will just have less to sell and the less they have to sell the less money they can bring in from customers. Now fast growth on the part of those who peer aggressively doesn't mean that you still can't get ahead of yourself by borrowing too much money.

COOK Report: OK. I always had the impression that the successful business model was supposed to be one of keeping as much traffic as local as you can by doing as much peering as you can. But that ultimately as you go upwards in the hierarchical tree, sooner or later you get to people like the Tier Ones who say that if you want to get to the rest of the internet your non peered traffic is x megabits per second for which we will charge you y dollars per megabit per second to deliver.

Woodcock: That is exactly what the function of a transit provider should be. Someone to charge you some price for traffic that you either cannot deliver or don't want to deliver yourself. Zocalo never had peering in the Far East because, although we had a large volume of traffic that went there, it was never quite large enough to justify pulling a DS3 across the Pacific. It would have been phenomenally expensive. It was much easier simply to buy transit at the PAIX and hand off to a transit provider that did have that kind of connectivity and let them worry about it.

This is the kind of decision about economic balance that every provider has to make. It was the fact that we were dumping off traffic everywhere that it made sense to do so that gave us the economic edge and allowed us to be profitable. If we were simply reselling transit the quality of our service would simply be no better than the service of the people from whom we were buying.

COOK Report: But what segment of the market was doing things as carefully as you?

Woodcock: Let me try to answer that for you by building a model of a minimalist ISP. Such a model will be a really useful building block in understanding what has happened with exchange points and how exchange points differentiate themselves.

The Model of the Minimalist ISP

The minimal ISP will have two transit

providers. It will have transit from two different upstreams and it will have peering. Let's try stripping this down and see why it doesn't work if you have anything less than this. If you just have peering you cannot sell transit to someone else because transit means access to the whole Internet. You cannot peer with the whole Internet. No one in fact can peer with the whole Internet. Someone is always in some sense your upstream. If you are just peering then, you don't have connectivity to the whole Internet and you have to buy transit from someone. So you have to have a transit provider. But if you have only a single transit provider, what happens when that transit provider goes down? As it most assuredly will. It will either be them or the tail circuit to them sooner or later. To fulfill your obligations to always provide transit, you need redundancy.

Now let's say you have two transit providers only and no peers. The reason why that isn't going to work is because the transit providers have a certain value that they are selling at a certain price. If you are reselling only that and adding your own costs, the cost of your pipes to your customers will be higher than theirs.

COOK Report: What if your two transit providers were UUNET and Sprint?

Woodcock: It probably doesn't change things. My point is that if you are buying transit from two providers and you are adding your costs and your profit margin and reselling the result, you price to customers is going to be higher than their price to their customers. And your customer might as well bypass you and go directly to them. You have no value add.

To be able to sell reliable transit, you have to have two providers. In order to be able to have a value added so that you can stay in business, you have to peer. Because delivering of traffic locally at a lower cost than transit will reduce the delivery cost per bit of your traffic from what your upstream transit providers are charging you.

COOK Report: So it is only what you can add to your balance sheets via peering that gives you an opportunity to overcome the burdens of your transit provider charges?

Woodcock: The more peering you have and have at lower cost the better your balance sheet. Consequently if you should decide to peer at a peering point that is really expensive – one with a high cost of entry and one where you can't move much traffic, you will wind up spending more than you save. You always want the cost of your peering points to be as cheap as possible.

There are distinctions between peering points. Some are expensive and some are cheap. Logically – if you are peering, you want to do so as cheaply as possible because the only thing that makes one peering point cost more than another is more reliability. But you don't want to pay for reliability at a peering point because a really unreliable peering point might be down for 45 minutes or an hour. What is wrong with a peering point being down for an hour? You don't ship traffic trough it for an hour. For that hour your average delivery cost per bit goes up slightly because you are shipping through transit for that hour only.

If you look at the report we did called introduction to peering economics, you will see a slide that shows what the cost of reliability is. Using very reasonable numbers it is easy to show you can spend $10,000 a month extra on reliability in order to save five dollars a month on additional transit. www/pch.net/documents/papers/intro-economics/intro-ep-economics.ppt

Basically there are many exchange points that cost (for a 100 megabit or a gigabit port) between $50 a month and 500 dollars per month.

COOK Report: The cost is for a port into a switching fabric? Once you have the port you establish BGP peering sessions with your peers.

Peering: Mechanics and Cost

Woodcock: Yes. The fabric most likely is just regular Ethernet. Your peer's routers are also likely on the Ethernet. All you are doing is just putting up BGP peering sessions across the Ethernet. Now the router is a one time cost and the space for the router should be pretty cheap. The line from your NOC to the exchange point is going to be your major recurring cost. It will be what will adds up over time. The exchange point itself should be cheap.

So let's surmise that people begin to think as much as they should about how the business they are in really works. Let's say they decide to spend more money in order to become more reliable. Spending this money will enable us to charge more. This will make us "better" than other exchange points. But in reality, the more costly the exchange point, the less economic sense it makes to be there.

COOK Report: Because the only reason that you are "there' is for this little amount that you can shave off your operating expenses? Right?

Woodcock: Exactly.

COOK Report: It sounds to me like Equinix fits into the model of more expensive than need be.

Woodcock: They've lowered their prices several times. When they got started doing business with them cost $7500 a month. Now it costs 700 a month. They are undoubtedly doing much better now than when they started. They are deriving more revenue from the rack space.

If you look at PAIX, you will find that it is not terribly dissimilar, but you will also find that PAIX's prices are higher. So here we have medium priced exchanges, although looking globally, the median price per month per a switched port at an exchange is probably 250 dollars.

COOK Report: What determines the

speed of the port?

Woodcock: What makes economic sense is for the port to be the fastest speed that is practical for all participants. Ports right now are mostly gigabit but some are also only 100 base T. No one wants to pay for a gigabit interface to a router if he can't ship more than a megabit per second. No one at this point is doing 10 megabits and only a few people are doing ten gigabits. Basically there are really only two port speeds that are in use by 99% of every one who peers – 100 megabits and one gigabit.

COOK Report: So if you are talking about an exchange point, a neutral exchange point is really the only kind worth talking about and furthermore one with more low tech than high tech technology is often imperative. You want security but you don't want fancy hi tech sophisticated security.

Woodcock: Yes. Remember that different people derive different amounts of value from an exchange. Also you must remember that no one wants to subsidize their competitor. Another big factor in exchange point economics counts very heavily. There is a need not just to minimize the total cost but rather to minimize the shared costs. You are much happier spending unshared money than shared. You are much happier spending on something when the full benefit accrues to you that you are when you must pay for the share of a combined cost.

COOK Report: Yes. What things can happen to tip things in one direction versus the other?

Woodcock: With something like Equinix, you have a big building and lots of staff and lots of handprint readers. This is all shared expense. At the other extreme look at the SIX, the Seattle Internet Exchange which has never charged dues ever. A gigabit port is free. The way this works is they bartered a broom closet (literally) from building management. They put a switch in the closet and avoid co location fees. There is no staff and no handprint reader. The broom closet is adjacent to the riser in the Westin building. As a result there is no shared cost. And no one has to trust anyone else. Not having to trust your competitor providers an interesting dynamic. Traditionally the exchange has been to get competitors talking and exchanging information and really working together. But there is a huge difference between trusting you competitor and HAVING to trust your competitor not to take advantage of a shared resource.

That's partly why Ed Kern's trick of putting a space heater in Mae West resonated so strongly with so many people. On the one hand they appreciated the humor of it and someone wanting to drive home the point to WorldCom that they needed better cooling and power. On the other hand this was dangerous in a way that affected all of them who would not have been affected had it not been a shared space with shared power and shared cooling. So this issue of trust is something that evolves well and is forced poorly yet it makes a very big difference in exchanges.

Having a low recurring cost and a low shared cost helps exchanges get going really well because participants can come in without having to already trust each other. Why because the cost of trying the exchange and the trust experiment and thus the risk to them is low.

This then really begs the question about the existence of the expensive exchanges. There are five to ten that cost upwards of $10,000 per month per participant. Why would anyone pay such a premium? Either they are phenomenally stupid or something else is going on here. Now the more costly exchanges are more into selling services to end user customers rather than to ISP who are much more used to constructing infrastructure for themselves.

COOK Report: What you are saying is that a place where ISPs can engage in cost effective peering with other ISPs is a very different beast than a place where you do commercial web hosting for end users.

Woodcock: Exactly. The answer is that, at places like Equinix and PAIX, people are doing transit. People are going there in order to buy transit from someone else. Now consider your minimal ISP with two transit providers and peering. Two transit providers because transit has to be reliable. Transit is why PAIX and Equinix are able to charge more than 250 bucks a month because reliability is not a service of peering. No one really cares about the reliability of peering. They only say they do.

The Tier Ones Don't Matter

This is going to sound kind of weird but when you get right down to it the Tier ones don't matter. They are insignificant in the over all scheme I things. There are too few of them. They aren't playing in the real economic space. Their dollars are not real dollars because they are being subsidized by other business units. They are not profit centers within their companies. They are not doing anything to lead the market in new directions.

COOK Report: so the point is that there are so many Tier 2 possibilities for routing traffic that Tier one stranglehold are in danger of being broken? Perhaps this is one reason why when E-bone was shut off a few weeks ago its traffic was absorbed with barely a hiccough?

Woodcock: Right. I am not saying that if the Tier one backbones were turned off there would not be a huge effect on the internet. There would be. But I am saying that in terms of guiding the development of the internet business model as a whole, they are not major players. They don't lead by example. Because their income comes from voice minutes, their Internet business plans don't even have to be self-sufficient.

COOK Report: They just do their own thing in a vacuum?

Woodcock: Yes. They are not really affecting what anyone else does. There is no exchange point in the world that would be adversely affected if the Tier ones disappeared tomorrow. They simply don't have a major affect on the peer-

ing infrastructure which is why they don't make a major difference to me.

What we are seeing in looking at exchange points is a differentiation between peering exchanges and transit exchanges. Peering exchanges are really cheap and quite large because there are two ways that they become worthwhile. Either they can be really cheap or have lots and lots of peering or preferably both. In a given region you want to go to one exchange and peer with as many people as possible. A second peering exchange in the same reason is actually not a good idea.

Why? Because either you don't talk to everyone or you pay twice as much because you have to go to the expense of connecting to a second exchange. You divide your resources and increase your costs. Therefore having a second peering exchange in the same region is a bad thing.

Now what happens if you have a second transit exchange? You get more redundancy. This is a good thing. How many participants do you need at a transit exchange? Three sellers and three buyers and you have an open market. More than this is nice but not worth paying a premium for.

COOK Report: Given Equinix's hi tech exchanges doesn't this say that something like web hosting is going to be the only thing that keeps them alive?

Woodcock: What I would say is that if there are three or more transit providers in each Equinix facility, and are each willing to sell transit to people and they form the equivalent of a market that is they compete with each other, then they can get other players who will locate to buy their transit there as opposed to just coming by to do peering.

COOK Report: So you are saying that if you are going to be an economically viable multi-homed ISP as opposed to an Earthlink or AOL, you need to have a well thought out road map for how you will do peering and a very different roadmap for how you will do you transit?

Where Are Things Headed? Recognizing a Transit Exchange

Woodcock: Indeed. So let's take this a step further and look at where this is going now? A new and very interesting problem today of down time when you purchase transit is the interruption that ensues when your provider goes out of business. Now it used to be that you could count on back hoes for 6 to 8 hours of down time per year, but when a provider goes out of business, you are talking days and maybe even weeks before you can get a new circuit.

The very best case might be 48 hours if your problem started long enough before the onset of a weekend. But even if you were multi-homed and suffered only the best case inconvenience, you would have to cope with restrained capacity for the whole time. You are at much higher risk because your redundancy is either diminished or gone.

There is now a new way of getting around this. You connect somewhere where you can actually walk over and plug into another ISP's infrastructure. This means locating in a transit exchange where there are three or more transit providers.

COOK Report: Is this what is meant by being able to buy transit at an exchange on the spur of the moment at prevailing market prices?

Woodcock: Yes. The spur of the moment is when the previous provider goes out of business and you really need it. You need to avoid being held hostage to a high price by someone who can take advantage of you because you are in a weak position. In order not to get taken advantage of, you have to be some place where there is an open and transparent market. If you are on the other end of a tail circuit, then you are locked in – one way or another they can take advantage of you and you are going to pay a premium. If you are at a transit exchange where they and their competitors are at they know that you are talking to their competitors also. They know they have to give you a fair price or that they will have no chance of getting your business.

But you still have the back hoe and the back hoe could take out your local loop between your pop, and the first transit exchange. So you do need to connect to a second exchange in the same general area as the first exchange. You don't want the second exchange to be further away because if it is, then you will have to pay a premium to reach it.

Now what we are saying is the new model says that you no longer need two transit providers and a peering point, you need two transit exchanges and a peering point because the providers themselves are getting too flakey.

Are we seeing much of this new model? Some. Not a huge amount yet, but definitely a bit. It certainly looks to be a trend. Although it may be a little premature to say with 100% certainty this is where things are going it certainly looks this way. I like this is as a direction and I think there are a fair number of people, many more now than there were six months ago who are liking it.

COOK Report: Pretty slick. But supposing I am an ISP and I need to figure all this out in a specific way that benefits me and not my competitors. How do I ascertain what a good transit exchange is? Does anyone keep lists?

Woodcock: There are a lot of people now who are willing to recognize this distinction but the distinction is also pretty new. So that if you go to an exchange point operator and ask the persons who are running the business end of the operation, and ask are you a peering exchange or a transit exchange they will say oh were are both of course!

COOK Report: Aren't we really talking about a subjective set of characteristics? As you look at possible points of interconnection for the time being you really have no choice but to have this imaginary mental map and say this exchanges

rules of operation and costs seems to push it in one direction or the other?

Woodcock: I think that there are places which are clearly peering exchanges. The SIX (Seattle exchange) is a classic example. No price and a huge number of participants. Everyone peers. No one trusts it enough to send transit across it. If it goes down, sure it will come back up the next time someone wakes up, realizes there is a problem and goes into the building to fix it. SIX then is clearly a peering exchange.

On the other side, there are exchanges that are clearly transit exchanges. For example the NAP of the Americas in Miami, Florida which is far too expensive for anyone to be able to justify peering there. I was working with one of their customers and looking at how much traffic they would have to push through it to justify the $35,000 a month they were spending there. It would have had to have been pushing over 10 gigabits of traffic through the exchange switch fabric for it to have been worth while but the switch fabric itself at this exchange is only a gigabit. The big value of the NAP of the Americas is as a market for global transit that sits right at the north end of a lot of Latin American fiber. Latin American ISPs can plug right into any of a dozen different transit providers, at market prices, with just a patch-cord to interconnect them.

COOK Report: But it was put together by Telcordia for Latin American phone companies to interconnect with North American phone companies?

Woodcock: Yes and although it may not be politically correct to say this, you and I both know that a Latin phone company hauling a circuit to Miami is not going to be getting free peering from Sprint. It will be paying for transit there. So that is clearly a transit exchange. $750 million dollars were put into the construction of that building.. Now compare that to a free broom closet and you will see the opposite ends of the spectrum.

So you might say that with these two extremes we have the two end points identified. But the problem is that everyone else comes some where in between. Unfortunately the model is not yet sufficiently clear that people are willing to hew to it.

COOK Report: How would you characterize 111 8th Ave and 60 Hudson Street in New York City?

Woodcock: There are people using both for both purposes. One thing that is becoming very clear to me is that it is difficult to take a facility and distinguish the facility as we did with the SIX and NAPof the Americas. Different people have different reasons for being there and different people have different costs of being there. So for one person for who it was really expensive the only reason would be to buy transit. Someone else may be borrowing a corner of someone's rack. For folk like that it could be almost free to be in an otherwise very expensive facility. Of course for some people the cost of being in a transit facility, where they must be, that they cannot afford even the minimal additional cost of being in a peering facility. For them there is no reason not to also get their peering done in the transit facility.

COOK Report: I suppose that with hindsight it is now all painfully obvious that for a long time few people gave such careful consideration to how to build networks with maximum cost effectiveness. But no they have no choice but to begin to think in the ways that you have been talking about.

Woodcock: Yes. And if you look at Asia, you will see that Asia never had enough money to do things wrong. That is one reason I really enjoy working over there. They think so much more clearly about this stuff. They haven't had the money but if you look at some of the things we have done they would not have replicated them even if they had had the money to do so. Why? Because they clearly don't make sense. People clearly were not thinking at the time when they spent a large amount of this money. Consequently a lot of people are winding up with these 'white elephants" that they are desperately trying to convert into something else or to advertise in some other way.

COOK Report: Who keeps an up to date list of exchanges?

Woodcock: Bill Manning, Antony Antony, and I were all maintaining separate lists, which we merged together about a year ago, and now jointly maintain. The merged list is at http://www.pch.net/documents/data/exchange-points/ It is a large excel spreadsheet listing about 300 exchanges globally. Of course there are new ones being started all the time. Ones that we do not know about. We try to bring new information to the listing task on a monthly basis.

COOK Report: If you are running your own ISP and it is large enough to merit multi homing then it seems to me that you must have information like this.

Necessity for Economic Modeling

Woodcock: I agree. Not only do you need the information, you need to think about what it means to you and you need to do some economic modeling . For example just because an exchange is there does not mean that you too need be there. Even if an exchange is clearly within your services area you don't need to be there unless you can save money by being there.

COOK Report: Who does the modeling? It sounds top me like there may be a service business here for someone.

Woodcock: Yes quite possibly. If you look at the modeling Stephen Stuart did for Abovenet and MFN, it parallels the modeling that I did for Zocalo and that Avi did for Abovenet. I think that a fair number of other ISPs, ones that were profitable and careful with their money did exactly the same thing.

What you do is to turn on Netflow in all your routers and you log the bit counts that are flowing through them to every other AS in the known universe. And you rank order your traffic. You see

where you traffic is going and you make a 'hit list'of the destinations. For example number one on such a list might be Sprint. Or more likely Cable and Wireless. Are either going to peer with you? Most likely not.

But there are two ways of looking at the list. The way that occurs to most people is to look at the adjacent peer. They say to whom do we connect right now that we are sending the most traffic to and how can we reduce that cost? The better way of looking at it takes more code, but a fair number of us do it this way. is to take the entire AS path between you and every destination and then you allocate points to each AS based on the amount of traffic that could potentially go through it. And then you must have a knob to help you determine how strongly you weight in favor of short paths. And you twist that knob back and forth a little bit and find that it changes your ranking. Suddenly you may find that Cable and Wireless, UUNET and Sprint are no longer among your top three. Pac Bell DSL, Chinanet and Road Runner might pop up at the top. What is significant is that there are ways to peer with folk like them because they are also interested in bypassing Sprint, Cable and Wireless and UUNET.

Your objective is to look at where your are sending traffic from the point of view of finding ways around those very large next hop destinations. It is not where you are sending traffic right now and who is already next to you. You want to reduce the amount of traffic that you are sending to places right now. The goal is how you can get around them and to reduce the amount of traffic going to them.

The question becomes whether there is an alternative way to get traffic to the final destination or more likely people a couple of hops upstream from the final destination but downstream enough from Cable and Wireless, UUNET and Sprint that they will be willing to peer with you because their interest will be the same as your interests. What it does is help you define yourself as part of the doughnut of (for lack of a better word) Tier Twos that surround the Tier Ones. And the doughnut is getting very fully meshed as the Tier Ones by peering only with themselves ensure that they become more and more isolated from the larger and larger portion of the world that is reachable from inside that doughnut.

COOK Report: What do the economics of doing business look like for people who are in the Tier 2 doughnut? I get the impression from talking with you that you are reasonably up-beat.

Woodcock: Yes. There is no question in my mind that you can profitably be either a local or a national ISP if you have a moderate amount of clue and you are really careful with your money. This is especially true if you have an opportunity to start fresh. One way or another. Whether you are coming out of bankruptcy or you are buying assets from companies that have gone under. This is in some ways a really good time to be in the business because a lot of people who don't know what they are doing and are way over funded are getting washed out of the business right now so you don't have to compete with these people anymore.

COOK Report: Is the fear of the huge AOL Time Warner Walled Garden model taking over the world now exaggerated?

Woodcock: They are not so huge anymore. They don't matter that much any more because they are growing at a less rapid rate than that of the Internet as a whole. Every day, they constitute a smaller portion of the whole. Secondly they have never made a dime. They can be a problem or you do a bit of judo and use their strength against them.

Here's the devious answer. You encourage all you customers to be multi-homed with a Sprint, a Cable and Wireless or a UUNET. Then you don't need any transit anymore, do you? You deliver the bits that are profitable for you to deliver. Everything else you let go through those transit providers. They don't know what it is that people are selling and don't know what it is either that people are buying from them. At some level they don't care because they are too worried about their auditors and not screwing up there.

We occasionally played those games when we were multi-homing our customers. We looked very carefully at what traffic we carried versus that which we pushed through the OTHER transit provider.

The Route Servers and Looking Glasses

COOK Report: Would you explain the mechanism of the Packet Clearing House Route Servers?

The way PCH got started is that we put up real route servers where everyone would actually peer with the route server. The route server would hand routes back and forth and then people would send the traffic directly to their neighbors. But the BGP sessions would be hub and spoke as it were rather than full mesh.

Think of the route server as a point source ISP which passes routes but not any bits. All it disseminates is routing. It says: Joe has a route to such and such and here is how you get there. John has a route to someplace else. Joe and John you guys go talk to each other about what you want to do with your routes. Just don't send me any bits. I am just handing off a message and no bits.

COOK Report: so it is like if I have 567 routes and Tom has 763 the server can show us the overlapping set of those that terminate at some specific point?

Woodcock: right. The route server can take everybody else's routes and aggregate them and hand you the best ones.

COOK Report: It is just like a highway map that shows you how to get from here to there?

Woodcock: Right. In the bad old days routers didn't have that much memory or CPU. Just pushing the bits took enough juice out of the router that for it to have to also do the routing was a kind of a pain in the posterior. You did not necessarily trust all the rest of your peers at the ex-

change to have a clue as to how to do this so you wanted their routes to be filtered before they got to you. Filtering after all really to a lot of CPU. So it made a lot of sense to separate the packet forwarding and the routing out into something like separate planes.

But now CPU and memory are not just that big a deal anymore. That function which made a huge amount of sense between 1994 and 1996 is not a really valuable one much of anywhere any more. For $1000 you can now get a Cisco router that holds 160 megs of RAM and is faster than anything they sold three years ago. So no one really needs this original router server function anymore but what is really necessary are looking glasses where someone can go and see what routes would be available to them if they went and located at an exchange.

What you need in conjunction with Netflow data about your own packets to simulate what your own participation at an exchange would have on your own traffic is looking glass data about routes available at an exchange should you agree to see up there.

You can also do a lot of other things with this information about routes if people are advertising it to you. You can compare what different people are seeing at different places and derive things from this comparison. For example in the last weekend of June we got some good numbers about just how big WorldCom really is by looking at what WorldCom is telling us about their routes and comparing it with what other people are telling us that Worldcom is telling them.

COOK Report: How does this compare with what Quarterman does at MatrixNet Systems?

Woodcock: I believe that he does a lot of ping and traceroutes rather than looking at actual routes rather than looking at OFFERED routes.. This will tell you something very different. It tells you how packets are actually flowing which is the net effect after policy has been applied to routes. There is a real difference between what people say they are going in their routing tables before you hand them the packet and what they actually do with the packet once you hand it over to them.

In some ways what Quarterman is measuring is more valid. The problem is that he is measuring real things after the fact on a small scale. The measurements are interesting but I do not believe that they are statistically significant. I maintain that the scale on which he works would have to be many hundreds of thousands of times larger before it can tell you.

COOK Report: How many of these looking glasses do you have to have at just exchanges in the US before you can offer sufficiently useful services to ISPs?

Woodcock: Our goal is to have one at every significant exchange. This means at about a third to a half of the 300 operational exchanges in the world. We have 13 or 14 in operation today.

COOK Report: Are they expensive? How difficult are they to put up?

Woodcock: We are a grant funded non profit. We do get good support from Cisco and some from Sun. The major expense for us is in actually going and doing the deployment work. Shipping the equipment, getting a person on site and hooking it all up. The hardware costs about $20,000 an exchange. The remaining costs of installation (travel and labor) are about 3 or 4 thousand per looking glass.

COOK Report: Do you see this ever becoming more like a cook book sort of thing? For example when Akamai wants to have an ISP to take one of its servers it presumably ships it with a plug in instruction annual or will it always remain much more difficult?

Woodcock: We have already done a lot of that kind of optimization. We are not just shipping things out because partly the function of PCH is to be a repository for knowledge about exchange points. It helps immensely to go to an exchange ahead of time. To put your feet on the ground to look around and take a few photographs. Understand who the people are and what the political issues are.

COOK Report: OK But if I am just average clueful ISPand want to get on your bandwagon. What must I do? I go to one of your seminars to find out what you have to offer.

Woodcock: We go to a lot of the ARIN, NANOG and APRICOT meetings. There we do one day tutorials on how exchanges work including how to participate in them and how to build them. We do a lot of one-on-one meetings with ISPs who want to try to do what they do better. Because we are grant funded we generally don't charge. Interested ISPs need to let us know and get on our schedule.

Where we are falling down a bit is in giving people tools to automate some of the discovery and the analysis. We are trying to build these tools and make then available.

COOK Report: Was CAIDA trying to build any such tools?

Woodcock: CAIDA has always worked a bit more closely with the carriers. They are much more interested in working with and for MCI and Sprint and Cable and Wireless as opposed to working with peers at exchange points. Their work is focused more on things like OC-192 Mon which is a bit of a fire hose for our needs. They are building very useful tools but they are focused at completely the opposite end of the spectrum.

By comparison if you look at the exchange we are helping to put up in Nepal most of the participants there are going to be connecting at 64 kilobits. But this interconnection infrastructure is critical for them. That 64 k will make the difference between them making money and them loosing money. It will make the difference a year from now between whether there will be two dozen ISPs in Nepal or two or three. It is an environment in which people have to be careful with their use of resources. I really like that.

COOK Report: To conclude: lets go back to the statements of Bernard Daines that perplexed me when I hear them in mid June. Have we just described in considerable detail how a service provider or an ISP might decouple the local loop and long haul requirements for offering remote broadband Internet services?

Model for a 'Global Scale' ISP

Woodcock: I think so. I think if you take for granted everything that I have just been saying, you will have some ideas about what it will mean to be a Global Scale ISPa year or two from now. What it means is that you do not have to be in the local loop business at all. What it means is that you need to touch down at a few transit exchanges in every country that you consider to be a significant market for your services. You don't do business with the phone company in that country. You get fiber from your landing point into that exchange. It is the exchange's responsibility to make sure it is available. In a lot of countries it will be the government's responsibility to make sure it is competitively priced. An enlightened government is going to want as many carriers as possible coming into its borders.

Individual carriers coming in from the outside cannot rip off that country's citizens. They want to encourage a market and discourage export of capital. There is a tutorial on the web site for policy folks in developing countries that explains the reverse subsidy problem. http://www.pch.net/documents/papers/policy-guide/ If you think about what this means it is that the next generation global carrier would touch down once in for example Singapore. They would have nothing to do with Singtel. Their customers would get circuits from their sites into the SOX (that's the Singapore exchange) and they would do an Ethernet cross connect with this 'new'global carrier. Or they would exchange traffic through the switch. You put in one router and one interface and there is no more physical configuration that needs to be done in that country until the link runs out of capacity.

It is really minimal in comparison with what you have to go through right now. There is almost no regulatory stuff that has to be done. There are no SONET Circuit level interconnects with another carrier that need to be made. The monetary investment is much smaller. This is why this model is going to make it much easier for ISPs to go global. A decent size ISPwill suddenly be able to become very big geographically. And this is its strength. If you want to be able to sell transit to someone, you want to be able to sell "haulage" to places where you customer needs to go but is never going to go itself.

COOK Report: So is the new business model then from a half dozen points in the US being able to by a lightwave to two dozen points on the globe?

Woodcock: If you were trying to become a global carrier, you would pick certainly no more than half a dozen transit exchanges in the US.

COOK Report: These are the new carriers then that emerge from the bankruptcies?

Woodcock: Yes. But admittedly predicting this is a bit of blue sky. Who knows if it will go this way? It is a picture I like pretty well. It is a picture that a fair number of folks who I consider smart also like.

COOK Report: As I have been looking at the edges of the network for the last few months I have certainly been wondering: how do you solve the long haul problem?

Woodcock: The only way to solve it is to decouple it from the local loop. The LECs are not going to give up their last mile monopolies. Period. Consequently you have to decouple yourself from them.

COOK Report: And what we have been talking about in this interview is one way to begin to do that?

Woodcock: Yes. You do it by distributing the work. The customer has to do a bit more of the work but he gets much more advantage. You do a little bit less of the work but you cover a bigger area.

COOK Report: Since there is a huge amount of fiber out there and the cost of lighting it with coarse wave division gig e has come way way down, folk find it possible to begin to experiment with their own build outs. Bit by bit, piece by piece they can begin to overbuild the structure that was there before.

Woodcock: And the complement to all this is that little local providers that are covering one metro area or one small country can go to their nearest exchange. Once there they can backhaul traffic for all the customers that are not large enough to take their own circuit into the exchange. You have a very synergistic complement between these two business models. What there isn't room for in this model is a regional. There isn't really any reason for someone to exist to exist who isn't either solving the global transit problem, or the
local-loop problem.

Part Two

Modeling of Netflow Data

COOK Report to Bill Woodcock on August 21: I understand that your major data source is Netflow output from Cisco or Juniper routers. Now Alex told me about a site in New Zealand http://www.gadgets.co.nz that makes its own Netflow based measurement tools. The proprietor is Martin van den Nieuwelaar (martin@gadgets.co.nz). Martin explains what he does in a very good essay for non technical folks. The essay http://www.gadgets.co.nz/software/traffic.shtml is called

"Traffic Traffic Everywhere - An analogy between network traffic and road users." There he concludes: "More recently, Cisco, in a gesture of kindness to network designers everywhere, began to implement what they call Netflow Exports.

These contain vast amounts of detail gathered from the IPpackets as they pass through a router. Such information allows the design of a more efficient network as they contain useful details such as source and destination IP addresses, Autonomous System numbers, UDP/TCP port numbers, as well as byte and packet counts."

Why Netflow

"With this information a road planner can see what destination countries the cars are driving to as well as the countries they will pass through on their way there. This is exactly the sort of detail the designer needs in order to efficiently choose the destination countries for the border teleportation units. The paths of drivers' journey can be shortened by minimizing the number of countries they must drive through (the number of hops as it is known), while also minimizing the number of cars that are sent via expensive super transport networks."

"Within the country the statistics also prove useful. The designer can now tell where cars are destined within the network. They can look at the cars on a particular road and see what their desired destination is. This is an important distinction. With SNMP statistics we can look at the cars traveling along a road from city X to city Y, but we know nothing about where they want to go. With Netflow Exports, we can see that most of the cars really aren't interested in city Y at all, but pass through to city Z. When we have this information, we have the power to make accurate design decisions. Depending on exactly how much of the traffic goes to city Z, we may well decide to build a road from city X to city Z, and either reduce or even entirely eliminate the road from city X to city Y. Drivers are happier now because they don't have to drive as far, and we have even saved some bandwidth as the city X to city Y road traffic is reduced."

"All of this may sound too good to be true. Indeed you might be wondering about the drawbacks of Netflow Exports. Indeed there are some. Firstly, they increase the load on a router. It is therefore important to make sure the routers have enough spare capacity before enabling Netflow Exports. The other problem is related to the volume of statistics. Potentially, a statistical record could be generated for each packet that flows through a router! This could lead to vast volumes of statistical data being produced. In reality things aren't this bad, with each record (called a flow) typically representing many packets. The volume of data is still significant however, and special techniques are used to reduce and manipulate the statistics. The vast quantity of statistics also means there are often numerous ways to interpret the resulting statistics, and for this reason special tools for their collection, analysis, and visualization have been built. Network Intelligence is such a product and its visualization engine is capable of displaying the traffic for an entire transit network. High speed three dimensional graphics based on OpenGL are used to immerse the user in a representation of their own IP network and associated traffic."

[**Editor's Note**: As we understand it Martin's software does path analysis. He did confirm on August 30 that his software did not do the Synthetic Path Analysis described below - we believe for the first time. Martin also pointed out that our just quoted excerpt from his Traffic Traffic Everywhere essay uses some terminology that is defined only in the early part of the essay that we don't quote. For those who are curious we do recommend using the URL to grab and read the entire essay.]

COOK Report to Bill **Woodcock** on August 21, 2002: Would you walk me through the data sent by Alex Tudor on the 19th? The first portion follows.

Tudor: An ISP needs to be continuously concerned with lowering its per bit cost. The following methodology (credited to Bill **Woodcock**) quantifies an ISP's usage of its resources, namely peering & transit connections and suggests possible changes. The numbers produced together with an ISP's actual cost can then be used for an actual monetary calculation.

Quantification of resource use simply means the amount of bi-directional traffic - volume in bytes - an ISP exchanges with its peers and transit providers. We differentiate between two traffic types: transit and terminating. Transit traffic passes through an AS; terminating traffic ends in an AS.

For the purpose of explaining the method, let us assume an ISP with one router and several interfaces connected to several transit providers and several exchange points. From the router we collect: a) per prefix aggregated flow volume (using Netflow) and b) a full routing table (RIB) snapshot.

We perform two sets of calculations, one using the selected AS path as reflected in the RIB (reflected in Tables One and Two), the other synthetically constructed (reflected in Table Three in two parts below). The meaning of 'synthetic AS path' is defined later. First, the 'real' AS path.

For each prefix we have multiple alternatives, depending on how many transit providers we have. For example prefix 10.0.0.0/8 may be offered by provider 1 via AS path '1 23 100' and provider 2 via AS path '2 6 100'. The router however will make a choice between the two offered paths and use it. This choice is reflected in the RIB for each prefix. Thus, for example, if our flow data shows 1000 bytes for prefix 10.0.0.0/8 and the chosen AS path was from provider 2, we will attribute 1000 bytes of 'transit volume' to each of AS 2 & 6 and 1000 bytes of 'terminating volume' to AS 100. The result of this daily calculation yields the following output. (See **Table One** bottom of next page.)

Woodcock: He is using the Netflow data which gives you bit counts for source and destination AS prefixs. You know how much data is going from you to any destination AS prefix which basically is one of about 112,000 IPsubnets to which there is a route advertised in the global routing table. You look at a route and say this prefix 128.10.10.0 /24 got two gigabits of traffic last month. The prefix

is a destination out there on the Internet. It is one external to your network. It could be an ISP router or a router at a large gov't or corporate network. What it is, other than the fact that it can be routed to, is not important.

You look at your routing table and there you see that this prefix is being announced by such and such and AS and the AS path is AS1, AS3856, AS 715. So AS 715 Zocalo.net is the originating AS and it is being transited by AS1 and AS3856. Now Zocalo buys transit from AS1 (BBN) which means that BBN must be buying transit from AS 3856. Then you have a choice. You know that you can peer directly with AS715 or you can peer with AS3856 or you can continue to buy transit from AS1.

Buying transit from AS1 gets you to the whole Internet and saves you form having to do any extra work. You just pay the money and your bits are delivered. Moreover if you peer with 3856, you will have to go somewhere to do that. But doing so is presumably less difficult than peering with AS715. What I am getting at is that the closer in AS hops an AS prefix is to you in the as path, the more aggregation has already happened. The gross example is that if you are a single homed customer of an ISP, one hop away from you, everything has been aggregated into one AS. Buying transit (ie. an Internet connection from that one AS) gets you the entire world. At the extreme other end, if you look at the right hand side of every AS path, you find all the different originating ASs (about 10,000 in all) through which traffic may be funneled to you.

COOK Report: Thus, depending on where you are in this branching structure, you have one funnel, two dozen, or many hundred funnels from which you may receive traffic.

Woodcock: But it is not as though you have a choice in having everything be one two or three hops away. Instead you get them in kind of 'rough buckets' of other people's choosing which are exchanges. At an exchange you will get some fairly random selection of AS's that you can peer with. These ASs will appear in different places in different AS paths for different prefixes.

The trick is that in looking at this rough bucket of different things, you know you have a cost to get to the exchange and a cost to participate at the exchange. And you have overhead associated with peering. So peering does incur a cost. But presumably there is also a savings because, if you don't peer, you have to spend money on transit. So the question is how many bits are we sending to or through ASs that peer at this exchange?

Study Where your Traffic Is Going

Now lets look at BBN, the first line on Table 1. What Alex presents in the first table is what actually happened. Total traffic flow for the 24 period of January 31, 2002. This is saying that 958 megabytes went through BBN of which 16 megabytes terminated within BBN. This means that BBN is a really good aggregator but not a very interesting destination. They are good to go through but not valuable to be able to reach in and of themselves. By the way the percentage figures 13.7 and .6 percent are the respective fractions of the day's total traffic for the entire network.

Lets go on to UUNET on the next line down. We see that UUNET transited slightly less traffic but terminated a lot more traffic. We know that for 5.4% of our traffic that terminated on UUNET that we have no other way of delivering it. This says that we either have to buy transit from UUNET in the amount of 144 megabytes per day or we have to buy transit from someone who peers with UUNET in the amount of 144 megabytes per day, or we have to peer with UUNET ourselves. But UUNET is not going to accept another peer because that would force the share of the donut hole to be split in more ways.

Table One:

Top 20 Organizations Transiting Traffic on January 31, 2002 (real)

	Organization Name	Transit		Terminating	
1	* BBN	(13.7%)	958 MB	(0.6%)	16 MB
2	* Uunet	(13.0%)	905 MB	(5.4%)	144 MB
3	* Sprint	(7.6%)	529 MB	(0.4%)	11 MB
4	* Cable & Wireless	(7.6%)	529 MB	(0.4%)	11 MB
5	AT&T	(4.4%)	307 MB	(5.4%)	144 MB
6	* Exodus	(2.7%)	186 MB	(6.5%)	174 MB
7	Qwest	(2.0%)	138 MB	(0.8%)	22 MB
8	Level3	(1.8%)	129 MB	(2.5%)	66 MB
9	Data Communications Bureau	(1.6%)	114 MB	(2.9%)	77 MB
10	EUNet/KPNQuest	(1.5%)	108 MB	(0.0%)	MB
11	Pacific Bell	(1.5%)	103 MB	(3.6%)	96 MB
12	*= Critical Path	(1.5%)	103 MB	(3.8%)	103 MB
13	= Cox Communications	(1.1%)	77 MB	(2.9%)	77 MB
14	Dante	(0.8%)	58 MB	(0.0%)	MB
15	France Telecom	(0.8%)	54 MB	(0.3%)	8 MB
16	* Abovenet	(0.8%)	53 MB	(0.7%)	18 MB
17	Verio	(0.7%)	51 MB	(1.5%)	41 MB
18	= University of London	(0.7%)	49 MB	(1.8%)	49 MB
19	Teleglobe	(0.7%)	49 MB	(0.0%)	MB
20	* CENIC Univ. of Calif.	(0.7%)	46 MB	(0.0%)	MB

'*' means that our ISP either peers or buys transit from the named organization.
'=' calls attention to the fact that terminating and transit traffic are equal. From this list one can tell, among other things, how one's transit providers are used, relative to each other.

Therefore they will either force us to buy 144 megabytes of transit with them per day or buy that amount from one of their peers. Since their peers occupy the rest of the donut hole they will agree and band together to enforce the oligopoly. Consequently, there is a certain amount of traffic there that we simply can't get away from.

COOK Report: How about entry 5: AT&T? There the percentage of the terminating traffic is actually greater than the percentage of sent traffic?

Woodcock. Correct. ATTas an ISP, has a large number of dial up users and they also terminate @Home. ATT would be easier to get peering with than UUNET because they need access to us (Zocalo) in the amount of 144 megabytes

COOK Report: It would be the same thing with Exodus I would think.

Woodcock. Yes Exodus is pretty trivial. Everyone peers with them. That is why they are as big as they are. It is because they have a lot to sell because they peer with a lot of folks.

COOK Report: They fit into your philosophy then?

Woodcock: Sure. All of these folks do. It is more a question of trend. Now let's look at number 11: Pacific Bell – 96 megs which is quite a lot. There is almost nothing else going on there. Out of 103 megs total, 96 terminates with them. Only seven megs goes on through them to someone else – namely their customers who also have ASs.

COOK Report: So on each line there are three things that are important. 1. The percentage of transiting traffic. 2. The percentage of terminating traffic, but also 3. The difference between the two traffic figures which, in its own right, is important to look at.

Woodcock: Correct. Now look at the use of notation in the table. The "*" says that there is a direct connection to that peer.

COOK Report: But the "*" says peering or transit. How do you know which is which?

Woodcock: What's the difference?

COOK Report: Money.

Woodcock: Exactly. But this table is not intended to show money. It is merely algorithmic. How does this know whether there is money? It cannot know that. I have access to my books and know from those other sources how to apply the 'money."

The equals sign meanwhile is saying that the transit and terminating amounts of traffic to a given AS prefix are identical. This means that these folk are really good to peer with directly. You get "full value" for peering with them directly. But this is also not to say that the ones with the equals signs are absolutely the best targets. What it saying is that if the equals folk are big, then there is no need to delve further in to the network topology. You can't improve on what you are getting by peering with a downstream customer of theirs.

BBN at 958 megabytes looks great. But if we look at the ratio between 958 and 16 megabytes, we see that yes we are delivering a lot of traffic through them. The reason for this is likely because of the star. The star means we are directly connected. They are offering a lot of desirable routes. But we could do a lot better, if our tools could help us find out whether we could find alternate means of delivery. Thus, in this table, being at the top does not indicate desirability. It indicates what actually happened.

The next table is Table Two on the bottom of this page. It is sorted by termination and says to whom we actually sent the most traffic. This is rather valuable.

Woodcock: You can look at this table

Table Two:

Top 20 Organizations Terminating Traffic on January 31 2002 (real)

Tudor: We can also show the calculation in decreasing order of terminating organizations, as shown below:

	Organization Name	Transit		Terminating	
1	* Exodus	(2.7%)	186 MB	(6.5%)	174 MB
2	AT&T	(4.4%)	307 MB	(5.4%)	144 MB
3	* Uunet	(13.0%)	905 MB	(5.4%)	144 MB
4	*= Critical Path	(1.5%)	103 MB	(3.8%)	103 MB
5	Pacific Bell	(1.5%)	103 MB	(3.6%)	96 MB
6	Data Communications Bureau	(1.6%)	114 MB	(2.9%)	77 MB
7	= Cox Communications	(1.1%)	77 MB	(2.9%)	77 MB
8	Level3	(1.8%)	129 MB	(2.5%)	66 MB
9	* Sprint	(8.3%)	580 MB	(2.1%)	56 MB
10	= University of London	(0.7%)	49 MB	(1.8%)	49 MB
11	= America Online	(0.6%)	42 MB	(1.6%)	42 MB
12	*= Hotmail Corporation	(0.6%)	41 MB	(1.5%)	41 MB
13	Verio	(0.7%)	51 MB	(1.5%)	41 MB
14	= Deutsche Forschurgsnetz	(0.5%)	38 MB	(1.4%)	38 MB
15	= Earthlink	(0.5%)	34 MB	(1.3%)	34 MB
16	Soutwestern Bell	(0.5%)	33 MB	(1.1%)	30 MB
17	ServiceCo - Road Runner	(0.5%)	32 MB	(1.1%)	30 MB
18	SURFNet	(0.4%)	30 MB	(1.1%)	29 MB
19	= Salesforce.com	(0.4%)	28 MB	(1.0%)	28 MB
20	= Cybercon	(0.4%)	25 MB	(0.9%)	25 MB

This list tells, for example, that Pacific Bell may be a good peering candidate.

(Table Two) and say we are already doing the one thing that would help us the most: peer with Exodus. We deliver 174 megabytes of traffic directly for which we do not pay. On the basis of the table not distinguishing between peering and transit, we could be buying. But I know that we are not buying and that this is a peering session. UUNETof course is transit. But AT&Tnumber two on the list is obviously the big target of opportunity. If we were to peer with them we would get 307 megabytes of traffic that went through someone with a star first before it got to them.

Now this implementation is over counting. Because if you sum up all of the transit you are going to come up with a huge number in comparison with what you get by summing up all the termination. The total for the termination column will add up to 100%. This should also be equal to the sum of the transit numbers from the lines that have asterisks or 'stars'. These are folks that are adjacent to us. The amount that all the folks who are adjacent to us transited of our traffic is the sum of all the traffic we sent. The sum of all the termination is also the sum of all the traffic we sent. Anyone directly connected to us is either a peer or source of transit. In Table Two Sprint and UUNET are transit providers while Exodus, Hotmail and Critical Path are peers.

Traffic to Hotmail was 41 megabytes. Hotmail doesn't have any customers other than stuff within their own network so all of the 41 megabytes we sent to them terminated right within their own network. If you sum up the transit numbers which is how much we sent that went through all the lines with asterisks you will get the total amount of bits that we sent out. Now we communicate with way more than 20 other AS system prefixes. Therefore some of our smaller peers (lines with asterisks) will not show up in these tables. If you look in the terminating column and add up for all AS system prefixes we terminated to, you will also find the total amount of traffic we sent. The transit numbers without an asterisk are therefore without direct connections to us. They are potential peers because we did in fact send traffic that had to go through them to get to its destination. But we were not peering with them so it also went through something else first. Now this means that they were a customer of someone else, whom we could bypass, if we went directly to them to peer. But it does not tell us either how difficult it would be or expensive it would be to actually go out and peer with them directly.

The University of London at 49 megabytes is number 10 on the list. If we wanted to peer with University of London, we'd probably find ourselves out of luck. The University likely simply does not peer. America on Line is number 11. AOL does peer. We would need to go to places where AOL peers, contact them, establish a peering agreement and then we would reduce by 42 megabytes from one of the starred numbers. We would then show an asterisk for AOL.

This is a preliminary analysis that does not take money into account. It just tells us what ASs are good targets for peering.

COOK Report: Southwestern Bell, Earthlink, and Road Runner both look like attractive targets.

Woodcock: Right. All are big broadband end user systems. Lots of eyeballs there.

COOK Report: Deutsche Forschurgsnetz fits the pattern but the problem with them might be you'd have to go to Europe to peer?

Woodcock: Right. And I don't recognize them moreover. The fact that I don't means that that traffic was probably short term burst of traffic from one of our customers and was not representative of a long-term trend.

COOK Report: Therefore, the next layer of complexity is that to be sure of this you have to do it a few times a week for at least a month. Right?

Woodcock: The way we did this was to run it every day. In addition we ran aggregates over a longer period of time. If you watch it on a daily basis, you will see a huge amount of churn with destinations moving up and down the ranks. If you watch it every day over a longer averaging period (for example a month) you should find that those results will stabilize a lot more. Cable and Wireless and those folk will always rise to the top.

To do the necessary planning one is going to look at daily, weekly and monthly snapshots. Looking at it daily lets you spot trends earlier. A trend will pop up and then appear with greater and greater frequency eventually working its way up into the top in a longer-term average. If I see Deutsche Forschurgsnetz twice this week, 4 times next week and everyday there after I can tell that something is happening even though they would not come up toward the top of the monthly average until the following month.

Now Alex is saying that Pac Bell is a good peering candidate because they are number five on the list and there is no asterisk there and it is more likely that they are going to want peer with us than AT&T is. Why? Look at the ratio: 103 to 96 as opposed to the 307 to 144 AT&T ratio. The closer the ratio is to equal, the more likely that party is to want to peer with us.

COOK Report: Now with the third table it is going to get really interesting.

Woodcock: Indeed Table 3 on the next page is a bit more complicated.

Woodcock: In these final two tables we have synthesized a composite AS path. This is saying that we are not looking at the AS paths for selected routes but that we are looking at ALL of the potential ones – everything that we are being offered.

In the first two tables we were only looking at the subset of the routing table that we were actually using. In the first one we were receiving routes from everyone that has an asterisk. But we are not necessarily using every route that we hear. We are picking based upon a BGP selection algorithm, one route to every destination. The first table used our actual

Table Three:
Top 20 Organizations Transiting Traffic for 20020131 (synthetic)

Tudor: How would the answer change if the router selected provider 1 [as a peer], in the example above? To answer the question I introduce the concept of 'synthetic' AS path, which in the example above is built by concatenating all ASes in the path and eliminating the duplicates, while retaining the destination AS (the last on in the path, e.g. AS 100 in the example above). Using the prefix 10.0.0.0/8 from above we would get '1 2 6 23 100'. We now run the same calculation as above, except that instead of using the router-selected path we use the 'synthetic' one. Below is an example of the result, run for the same day as the output above.

	Organization Name	Transit		Terminating	
1 *	Cable & Wireless	(18.5%)	3351 MB	(0.4%)	11 MB
2 *	Sprint	(17.8%)	3223 MB	(2.1%)	56 MB
3 *	Uunet	(15.9%)	2869 MB	(5.4%)	144 MB
4 *	BBN	(13.9%)	2517 MB	(0.6%)	16 MB
5 *	ViaNet Communications	(2.9%)	532 MB	(0.0%)	MB
6	AT&T	(2.2%)	403 MB	(5.4%)	144 MB
7	Qwest	(2.1%)	387 MB	(0.8%)	22 MB
8 *	Abovenet	(1.8%)	325 MB	(0.7%)	18 MB
9 *	Exodus	(1.4%)	245 MB	(6.5%)	174 MB
10 *	Hurricane Electric	(1.1%)	205 MB	(0.6%)	16 MB
11	Level3	(1.1%)	196 MB	(2.5%)	66 MB
12	Teleglobe	(0.9%)	162 MB	(0.0%)	MB
13	Data Communications Bureau	(0.6%)	114 MB	(2.9%)	77 MB
14	Pacific Bell	(0.6%)	112 MB	(3.6%)	96 MB
15	EUNet/KPNQuest	(0.6%)	111 MB	(0.0%)	MB
16 *=	Critical Path	(0.6%)	103 MB	(3.8%)	103 MB
17 *	Cogent Communications	(0.5%)	97 MB	(0.0%)	MB
18	Global Crossing	(0.5%)	86 MB	(0.2%)	5 MB
19	Verio	(0.5%)	82 MB	(1.5%)	40 MB
20 *	CENIC Univ. of Calif.	(0.4%)	77 MB	(0.0%)	MB

This ordering, by transiting volume, suggests optimal (for the traffic mix of our ISP) transit providers. By ordering the list by terminating volume we get a suggested list of peers, as below.

Top 20 Organizations Terminating Traffic for 20020131 (synthetic)

	Organization Name	Transit		Terminating	
1 *	Exodus	(1.4%)	245 MB	(6.5%)	174 MB
2	AT&T	(2.2%)	403 MB	(5.4%)	144 MB
3 *	Uunet	(15.9%)	2869 MB	(5.4%)	144 MB
4 *=	Critical Path	(0.6%)	103 MB	(3.8%)	103 MB
5	Pacific Bell	(0.6%)	112 MB	(3.6%)	96 MB
6	Data Communications Bureau	(0.6%)	114 MB	(2.9%)	77 MB
7	= Cox Communications	(0.4%)	77 MB	(2.9%)	77 MB
8	Level3	(1.1%)	196 MB	(2.5%)	66 MB
9 *	Sprint	(17.8%)	3223 MB	(2.1%)	56 MB
10	= University of London	(0.3%)	49 MB	(1.8%)	49 MB
11	= America Online	(0.2%)	42 MB	(1.6%)	42 MB
12 *=	Hotmail Corporation	(0.2%)	41 MB	(1.5%)	41 MB
13	Verio	(0.5%)	82 MB	(1.5%)	40 MB
14	= Deutschef Forschurgsnetz	(0.2%)	38 MB	(1.4%)	38 MB
15	= Earthlink	(0.2%)	34 MB	(1.3%)	34 MB
16	ServiceCo - Road Runner	(0.2%)	33 MB	(1.1%)	30 MB
17	Soutwestern Bell	(0.2%)	33 MB	(1.1%)	30 MB
18	SURFNet	(0.2%)	30 MB	(1.1%)	29 MB
19	= Salesforce.com	(0.2%)	28 MB	(1.0%)	28 MB
20	Cybercon	(0.1%)	27 MB	(0.9%)	25 MB

Often the 'real' and 'synthetic' rankings coincide. That means that you are buying transit from the right company and that your are peering with the 'right' ISPs.

routes. But it is neither as thorough nor as complete as it could be, because we have additional information. If our BGP selection algorithm were more clever, it could hypothetically deliver traffic in different ways to different peers. There are a lot of other criteria it could be using. Or it could be apply the same criteria, but in a different order to get a different result. What we are doing here is saying that we were offered five different ways of getting to this destination.

Instead of just looking at the potential of peering with people who happen to be in the path that we actually used, as a way of shortening the path, what if we peered with someone who was in the middle of the path that we were not using before? Obviously it would not be a win for this particular destination. But if we peered with that entity, it might be that doing so would bring us closer to a whole lot of destinations for each of which we have some traffic. This is saying that while someone might not be the greatest peer in any particular case, it might be that peering with them would bring us a little closer o a lot of the Internet as opposed to a lot closer to some specific part of the Internet.

The real BGP decision-making algorithm that is applied here is myopic. It looks at one prefix at a time. It asks what is the shortest path (fewest AS outer hops) to this particular prefix. It ignores the amount of traffic. It ignores dollar costs. It ignores a lot of things. One of the things it ignores is how close and AS is to other destinations that it doesn't care about for delivery of a specific packet. BGP delivers by whatever shortest path it has been given. It doesn't care that a packet may be deliverable via a route other than the one it has been given, it follows only the route it knows about, because this analysis is not looking at the specific path chosen by the BGPselection algorithm. It is looking at all the paths. And it is creating a synthetic path that contains in an unordered list everyone of the ASs that could be between us and the destination. Not the ASs that were actually between us and the destination in the path that was historically chosen by the router at the time the packet was delivered.

COOK Report: It is as though the AAA actually sends you from New Jersey via the scenic route or the god-knows-what route.

Woodcock: Let's see. New Jersey to Yosemite. We look at map with a specific destination in mind and we figure out what road to take. Say our choice takes us through Chicago and Denver. We look at the way we usually travel and say that indeed Chicago and Denver are a pretty good way of getting to Yosemite. In point of fact, if we had started in Denver it would have been pretty short. And if we had started in Chicago, it would have been shorter than starting in NJ. And this is good for knowing about getting to Yosemite but what about getting to Dallas. If we started in Denver how much help would that be in getting to Dallas? If we started in Chicago how much help would that be? In the real world we have traffic for both Yosemite and Dallas. What we want is a medium. We want some way of getting to Yosemite that will also bring us closer to Dallas and such a way might be though Omaha. So the analysis in Table 3 is telling us whether there are Omahas out there. Are there places that ma\y not have looked good for anyone specific destination but were really happy mediums for multiple destinations. They were less bad for all the others.

COOK Report: Is there an analogy here between what you are saying and what is necessary to do to select a good site for an airline hub?

Woodcock: Yes. Kansas City might be a great airline hub although it's horrible as a destination. It is not really close to anything. But as a take off point for a lot of other destinations it is not really bad. And if you look at how 'not bad'it is, the not badness far outweighs the lack of specific goodness.

Now look at Table Three and remember the over counting I mentioned. Look at how much more this is over counting. Why? Because this is all the times we could have gone through something and did not in addition to the ways we could have and did. The one before was could have and did. This is could have and did or did not. This is all the possible ways of getting from one point to another aggregated together.

It only counts each AS once per bit, but many more bits could go through an AS than actually did. In some ways this table is more interesting and in some ways it is less interesting. This tells us that UUNET, Sprint and Cable and Wireless have really good connectivity. And BBN has pretty good connectivity. The reason for this good connectivity is that all of them could deliver traffic to anyone because those guys all peer with each other and everybody is "downstream" from one of them. When you go downward in the table from these there is a huge jump. Now why is AT&T not up in the 10 to 12 range? It is not because there is no star in front of the entry and no star means that we are not directly connected to ATT. Since we are not directly connected to them, we only see the routes for them that we get through someone else.

COOK Report: If you were directly connected to Level 3, then the 196 and 66 megabyte figures would be a lot higher?

Woodcock: yes. Because we would be seeing more routes from Level 3. Anything we see in Table three from someone with no asterisk has been already "predigested" by someone who does have an asterisk. We had to have learned these routes from somewhere. Some one has already gone through all of AT&Ts routes and thrown away the one they don't like before they feed us the remainder. Only the guys with asterisks in this comparison are being validly compared. You may ask whether it is reasonable to say that they have been fairly measured.

Let's look at Number 5: ViaNet Communications at 2.9% which is way down from the 13.9% of the entry above it in the table. That is a fair comparison because it has an asterisk.

COOK Report: It must be peering?

Woodcock: No in point of fact we are buying transit from them and by means of the data in this table we see that their connectivity in comparison to our four main transit providers leaves much to be desired. Why are they above AboveNet and Exodus? Because Abovenet and Exodus are peering, we are only getting customer routes from them. Again to evaluate the data in the table one has to know where the dollars are going in order to know what is transit and what is peering. Or, one could look at the size of the routing table that was being received from each and that would help tell you as well.

Now how do you take care of the problem created by the unfair comparison? You go and you get the other routes from these folks and you throw them into the mix. You take a static dump from one of their customers and this algorithm will then fairly compare them. This is why looking glasses that receive full routes from peers are useful.

Why Looking Glasses Are Important

You can use a looking glass that is getting full route views to ascertain how valuable transit from that entity would be instead of just telling you how valuable peering would be. This is why when I go out to an exchange point, I put two looking glasses there. One of them asks that people give transit to. The other asks that people send only their peering routes to it.

To amplify this topic – Packet Clearing House puts out two looking glasses at every exchange. One of them we ask people to peer with – just as they would do with any real peer. In that looking glass you can see who other peoples customers are. You can find out what the value of peeing with them would be, because it peers with them and if you peered with them you would see the same things that it sees. I ask people to give the other looking glass transit. That is to treat it as though it were buying transit. By looking in that looking glass you can find out what you would see if you bought transit from them.

When you buy transit, you know you will get about 110,000 routes. What you don't know is how attractive those routes would be from your network's point of view to those that you would get from say UUNET or BBN.

COOK Report: In other words if everyone dumps transit routes into the looking glass, you can compare them and see from the point of view of your own traffic which offers the best match.

Woodcock: All you need is a one-time dump of that routing table in order to be able to flow your traffic tough it in a model. You can see where your traffic would have gone. Routes are routes. You dump them in together and reflow your traffic through them and you will see where your traffic would have gone. That is the elegance of this synthesized comparison. You can scrape up data of the street and through it in regardless of where it came from. Anyone's point of view is valid if you throw it into the synthetic comparison because it doesn't care too much about position in the AS path. It mostly cares whether you are in the AS path. You can refine this by weighting every single AS for every prefix by its distance from the right hand side. This tells you how good the route would have been if it had started from that point in the AS.

COOK Report: Good relative to what?

Woodcock: Its length. If we have an AS path that is five hops long, we say that this has a goodness of five where the smaller number of hops is better. If we ask what would happen if we peered with the guy in the middle of that, we would be able to lop two off and we would have a "goodness" of three – a considerable improvement.

This is not measured from the left hand side (us as the sender of the traffic) but from the right hand side which we define as the origin of the AS prefix which we need to be able to reach – ie to get packets to. We know then if we have something that is three from the right, it is better than something that is four from the right. The origin is the AS that is handing out a route to a destination and asking for traffic to be delivered to it via their AS. Now to get traffic to where it needs to go, it must pass through multiple ASs and the closer any AS is to the AS of where the traffic needs to go the better it is to peer with that AS.

This is a further refinement of deciding between Kansas City and Omaha because it tells you for all of the destinations that you were going to but were getting there by some other route how much better for example Kansas City would be than Omaha or St Louis.

These are two refinements one of which is possible with the code that Alex wrote but was not done in this instance – namely throwing in additional data that he scraped up off the street. In actual fact all of the data you looked at that Alex supplied in the four tables came out of one router. This is not the sum of data taken from a whole lot of different routers. And Zocalo does have multiple routers.

You take only your own Netflow data, not that from someone else, because such would give you wrong answers. But you can take their routes and get information from them that you need to engineer your own network. They tell you more ways that you could have gotten to places. They have more adjacencies that are different from yours. All of that is data that you want. You are looking for potential routes (roads) for your traffic to travel on. Our routes are the only roads that we can see if we do not do this kind of planning. If we look at someone else's routes (road map) the combined road map will show us more ways that we could get traffic where we need it to go. It gives us a better model.

COOK Report: What it sounds like you are saying is that what Alex compiled and sent is just a kind of single slice over view. And, if you are really doing this in a serious way, because the bottom line is whether or not you can run at a profit, you would be getting a lot more data and many more measurements on which to base your peering and transit decisions.

Woodcock: Yes.

COOK Report: Then doesn't it also follow that once you work all this out and take it as a tool by means of which you will formulate many of your business decisions, you likely can get the various necessary measurements and calculations done in such a way that it becomes rather well automated. When you get it up and running it becomes a system that you can watch.

Woodcock: Yes and this is a part of what I am trying to do. Namely I am trying to get data collectors out there for people to use in doing this analysis. I am trying to get these looking glasses out there and peering with as many people as possible so that some of those asterisks will be as close to everything as possible. I want to see a lot of asterisks in people's tables assuming that they use Cisco's code base or a similar notation.

COOK Report: Are you saying that it is only with the presence of a looking glass that you can do the truly useful economic modeling?

Woodcock: If you only had access to information from either the folk whom you were actually peering with or actually buying transit from, you could do quite a bit of optimization. But there would also be huge parts of the world that you would not have a road map for and about which you would, as a result, know nothing. You need someone else point of view to see things about which you have no data and which therefore are invisible to you. This particular router happened to be buying transit from Cable and Wireless, Sprint, UUNET, and BBN. Those four transit purchases gave it a really good view of things. Most ISPs would not do as well from a single router view because they would not have four transit providers piped intro a single router. That is why this router was chosen because from one location it did have a pretty good view.

Benefits of the Methodology

COOK Report: But this methodology

can be applied with some benefit by almost any ISP anywhere? Some benefit being that an ISP might get say 10% of the cost savings that might be obtainable with a complete set of tools and data. Your total economic benefit attained would be a function of how thorough you were able to be in your acquisition of data and use of tools?

Woodcock: Yes. There is a lot of analysis and work that needs to be done. If you do the analysis you should then see what work you could do to optimize your network.

COOK Report: And if everyone starts doing this, you do see a much more sustainable stable economic mesh of connectivity that can be built?

Woodcock: Yes. Absolutely. The problem is that right now so many people are operating so naively. The number of ISPs who are not operating by the seat of their pants you can count on the fingers of two hands. Compared to the thousand and thousands of ISPs in existence this is not an impressive ratio. Maybe you need mutant hands. There are about a dozen people who are actually doing this type of analysis. Stephen Stuart from PAIX is one of that dozen. It would be interesting if it were possible to compare and contrast the approach I have outlined with the approach he would recommend.

But remember that I was just talking about two things that could be improved upon. One is more input from the looking glasses. The other is that weighting by distance from the right hand side. Alex's coding never got around to adding that. This is really important. Then all of what we have been talking about is just relative. I doesn't give you actual dollar figures. We did our dollar figure calculations by hand based up this framework. Additional code should be written to do the dollar figure analysis.

This code would take cost spreadsheets as input. You would have spreadsheets both for the costs of circuits and for exchange point costs. This is something that, to the best of my knowledge, no one has automated yet.

A Note to Our Readers

The article by Farooq Hussain is unchanged from November's issue as are the interviews with Bill Woodcock. The introduction pp. 1-5 is mostly new. All material from page 28 through 89 is new. New material has been added to the ICANN coverage that runs from 90 to 99. The RIAA article is unchanged from November. Finally the Extended Excerpts and Executive Summary (pp. 105 -118) did not appear in the November issue.

A Mini Encyclopedia of the Economics, Politics and Technology of Internet Interconnections

Our Experts Discuss Architecture, Traffic Flows, Transit and Bandwidth Costs, as Well as Market Economics

Participant's Introductions Highlights

Bill Woodcock

(August 3) I'm the Research Director at Packet Clearing House. PCH has been building and supporting exchanges since 1993. We do a lot of tutorials on IX construction. Three in Nairobi next week, two in Kampala the week after that, Singapore the week after that, Kathmandu the last week of August, Kita Kyushu the first week of September, Rhodes the second week of September, et cetera. We deploy and maintain data collection equipment at a lot of exchanges (about two dozen at the moment, adding another dozen by year-end, and probably another two dozen next year) and archive the results. Those routing data in the archive are used by academic researchers, who don't have any way of doing data collection directly themselves. We do a little bit of analysis ourselves, usually to highlight the possibilities of the data, and get other people started on more in-depth projects. A couple of weeks ago when the WorldCom bankruptcy story broke, CNN called us to find out what effect a UUNet shutdown would have on Internet connectivity, and we did a bit of number-crunching around which they based their first news segment.

We also have some rather lackadaisical tools-development projects, mostly aimed at creating analysis tools which peering coordinators can use to determine what exchanges they need to be present at, and who they need to establish peering sessions with. (More accurately, the analysis is generally of the economic tradeoff associated with different routes for the monthly traffic loads to each destination network.) And we've got some other projects that are just aimed at providing operational support to IXes and the peering community generally. The work we're doing with the Nepalese ISP industry association at the end of the month is an example of that, as is the similar work we did in Stockholm this spring, helping to work out the political, economic, and technical problems which were driving their exchange into insolvency and were causing ISPs to abandon it. Another example is our Inter-NOC Dial-By-ASN hotline phone project, where we're deploying VoIP hotline phones to all the major IX and backbone NOCs that want to participate, so they can ring each other directly, just by dialing an AS number.

Prior to working full-time with PCH, I founded and ran a leased-line regional ISPserving university and corporate customers in the United States, called Zocalo. I started it in 1989, and finally left the last of my responsibilities there at the end of 2001, although I'd been gone in all but name since mid-2000. I built it up from 9600 bps leased lines out of Berkeley and Santa Cruz, at its start, to a network with POPs all over the west coast, western US, and a few in the northeast. But I was always too tempted to use it as a testbed for research work, which wasn't always an efficient use of the business' resources, which is why I'm just focusing on research work now, and leave Zocalo to business folks.

I've been an active and regular IETF participant since 1994 or so, likewise ISOC, NANOG, RIPE, APRICOT/APNIC, APIA and ARIN. More recently NorduNOG and a bunch of the newer regional operations groups. I'm on the boards of directors of quite a few exchanges now, and the technical advisory boards of several others. If I don't watch out, board meetings will begin to compete with real work, for time on my schedule.

I spend a fair amount of time thinking about, and talking with people about, peering and transit economics. I live in Berkeley, California, just one city block away from the University. Actually, I maintain a residence there, but mostly live on airplanes. :-)

Gordon asked: what would you do to make life in the industry sustainable again? My answer: Beat anyone who even thought of the phrase "cross-subsidy" senseless. Sorry, that perhaps overstates it a bit.

But I guess I've just seen a lot of really good, smart, hard-working people with really good ideas and companies either driven out of business or rendered really cynical by idiots with no clue what they were doing and a ready supply of cash to allow them to destroy the market for Internet services. That's a sad thing. Fortunately it's mostly only happened in the US and to a lesser degree in western Europe. I'd love to see market prices for Internet services eventually stabilize somewhere above costs, but it's not clear how to get there from here, other than through a long and painful process of laundering debt off of useful resources through chains of bankruptcies and acquisitions.

Alex Tudor

Hello everyone! I have spent most of my professional career as a software developer for Hewlett-Packard where I have been involved in OS and network management products (Openview). A few years ago, as Agilent split from HP, I took a research job at Agilent Labs in Palo Alto. As some of you may know Agilent took over semiconductors, optics, OSS and testing systems from the old HP. The company is now focused on biotech and communications.

As I was searching for research ideas I came across a little known (to me) element of the Internet, known as exchange points. From the beginning I had the good fortune of running into Bill Woodcock, who has patiently guided my inquiries since. My first project was to help with Zocalo's traffic analysis.

The results of these projects suggested the need for an on-demand transit exchange (in the sense that a connecting ISP could add/drop transit providers within the hour or less). This seemed like an ideal vehicle for deploying a research test-bed and getting Agilent into a service business (as opposed to its traditional box business). With the technical problems nearly solved, I embarked on a building a business case and plan. Even though I had a good location, one interested tier-one and a few tier-twos, the idea died. Maybe it makes more sense now.

Setting aside business model considerations the main challenge is how to do it without putting futher strain on the global routing table and involved routers; is BGP4 the best protocol to deliver this function? As a research project (at least in the beginning) this was supposed to attempt to address these issues. The initial test-bed consisted of a route-server (based on modified Zebra) which is fed periodic changes from a routing registry according to contracts. Sellers have to trust the route-server's routing decisions, as derived from the contract system/registry in order to prevent router configuration changes during contract execution. The interesting, possibly proprietary, work would have come a bit later by trying to develop and automate a buyer's transit buying strategy.

Since the demise of that effort I focused on BGP. The scope of this work is to characterize normalcy (if such a thing exists) and to develop tools for the multi-homed enterprises, ISPs and exchange points. To that end I am working with RIPE's RIS group using the data from their Zebra collectors. A few years ago Craig Leibowitz and Abha Ahuja studied BGP and uncovered and publicized a number of issues. In the spirit of learning more about the protocol I set out to duplicate some of their work, specifically duplicate announcements, flaps and inter-announcement attribute change timings for the period of 8/1/01 to 1/1/02 and daily since 6/1/02. The work is only preliminary and I have not drawn any firm conclusions. Early results indicate that some problems persist. At this point my goal is to better understand if what I see constitutes a problem and if so to try to identify the cause. I believe that this study (and others, e.g. CAIDA) can show the need for certain tools to use in problem detection and isolation and policing of the 'routing table commons' (to paraphrase Geoff Houston).

Francois Menard

I'm Francois Menard and work as a project manager for IMS Experts-Conseils, a large Canadian ISO 9001 consulting engineering organization of about 130 employees and 4 offices. Since Dec 2000, I've done more than 30 projects/ studies on private fiber optic deployment in Canada for several school boards, universities and municipalities. Today we have about 33% of all schools and nearly all university and colleges on Dark Fibre in Quebec (more than 3000 establishments). Lots of these organizations are going through the RISQ network in Montreal for peering with the Internet. There is a fascinating amount of consolidation happening and I'm interested of understanding its effects on the economics of providing educational, commercial and residential next-generation services.

My initial objectives are to assess the implications of the venue of community networks and the improved business cases for independent ISP's through broadband open access and private Wi-Fi build outs on the global peering practices of tier-one ISP's.

Pedro Ferreira

I am currently a PhD student on transit vs. peering under the supervision of Marvin Sirbu, at CMU, and also Terry McGarty, at the MIT's program on Internet and Telecoms Convergence. I've just finished a double master in Computer Science and Technology Policy at MIT, in this subject, under the supervision of David Clark. My thesis is available at http://itc.mit.edu/itel/students/papers/ferreira_thesis.pdf.

As you can see, this is just a very preliminary study of transit vs. private peering, done upon the scarce resources (I mean data) that I had at the time. I am currently working on improving the model presented in this thesis to take into account the fact that most IBPs require peering at more than one location. Still, data is scarce, mostly in terms of accurate architectures for interconnection and routing equipment prices. I have extensive microeconomic and econometric skills and I can contribute at the level of dealing with data for analysis, if this exists. In any case, the model developed gives some very clear ideas about how to decide between peering and transit. This is part of my early contribution to the discussion. I can also forward a (smaller) paper version of this document, along with presentation slides.

So, my primary goal is to learn more about peering vs. transit and in different contexts. As Gordon mentioned, I am currently in Portugal, my home country, looking for data and information on peering in Western Europe. I have some knowledge about what is going on around Europe and we know that things are quite different from the US, where the eyeballs are, so I think it would be important to keep in mind that different solutions certainly apply to different countries. My ultimate goal is to include a chapter on my dissertation on analyzing peering vs. transit in a European country (eg. Portugal if I can get enough information and data to do so). I am also very interested in both the technological issues and the economic issues related to interconnection. For the former, I am particular interested in understanding the impact of multi-homing and load-balance, multicasting, etc... For the latter, I am very interested in discussing QoS at IXs and associated pricing strategies.

I share Bill Woodcock's view of "stable prices for Internet Services greater than costs" but I also share the idea that it is not at all easy to get there. There is too much variability on what Internet Services mean and, again, contexts are very different. Additionally, costs are known, but pricing models, in my opinion, lack analysis.

Roxane Googin

My name is Roxane and I am your worst nightmare. I care about asset allocation and how technology changes impact the overall economic scene. For about 2 years now I have been of the belief that IP technology represents a paradox. It is both our pathway to the future and economic kryptonite. It is so good, it is unfundable. You can read about my thinking in "The Paradox of the Perfect Network." See the Isenberg/Weinberger write-up at <http://www.netparadox.com/>

Bottom line, IP is such a perfect commodity, it guarantees its suppliers a loss on operations. For its job, of communications, this makes it perfect, because it neither imposes any preparatory restrictions on communications nor does it run out of capacity. Thus, the maximum creativity at the edges means zero value-add for the middle. Think of it as preserving entropy or something. This also means that no one makes money in the middle. This means our future well being is dependent upon something that no one benefits from doing.

It also means that all telcos on the planet go broke, taking their investors with them. This is where it gets messy. How do we get out of our legacy investments? Who pays for the next build-out? (Hint: it has to be social as the markets are not about to touch this one) This is a non-trivial question. The entire market decline is based on this problem. The crooks, the bankruptcies, the lack of revenues all stem from this one truth. The legacy telcos have about $1T in debt globally. It is worthless. Those bond holders will never get paid back. When do we admit this and quit waiting for the "bottom" or the "second half rebound"? There is not one. We just go down until we really crash. This is just a warm-up!

Then, how do we convince our government that the telcos are in fact the bad guys, and the crooks who stole money actually run the right stuff? For IPto win, someone other than the legacy Telco guys must manage the network. Its build-out cost must be sold to an increasingly impoverished public. The only guys who have not embezzled money are the ones who cannot be allowed to survive, because they will do anything they can to prove IP is useless and SONET is all there is. If SONET lives, our economy dies, as we cannot move to the next productivity paradigm of the real-time organization using Web services on anything other that ubiquitous, Gb Ethernet. If we have that, the legacy telcos are broke. It is that simple.

My bottom line question is: how much will a "good enough" buildout cost? How can we sell this to the Government as simple enough to manage they are not creating another postal service or Amtrak? How do we convince them that the cable high-speed buildout is a dangerous precedent because it merges transport with content, thus compromising free speech (we have been there before). How do we convince Government that their Telco friends are actually their worst enemy, and those scumbags who stole the money actually run the right networks?

Phil Weller

I currently serve as the CTO at FASTNET. I joined FASTNET in 1996 to build and run the engineering efforts. Prior to that, I spent sixteen years at AT&T, mostly in the Bell Labs side, but the last few years with the Microelectrons Buisiness Unit (now spun off as Agere) Global Datacenter.

I'm arguably the dark horse of this group. I have neither the academic background, the reputation, nor the focus of many of this group's fine participants. In that respect, one of my goals is to basically shut up and listen :).

What I do have is real life experience in keeping alive what many here (including Bill) might consider an endangered species - a near 100% transit regional ISP. So I'll bring a strong balance into the mix. Note I am not stating that I religiously believe in the 100% transit model. FASTNET's current network architecture is as much a factor of the founders'original vision and current economic times, as it is good engineering. FASTNET started in Bethlehem PA, which for some reason, was overlooked by all the players when building peering points :). Within the past few months, FASTNET merged with AppliedTheory and Avi's NetAxs, bringing in an overlap of peering at the MAEs. I'm betting discussions we will have on this list will mirror the exact discussions occurring inside FASTNET.

Pedro, Bill - thanks for making your research/viewpoints available. I look forward to relaxing with them over the weekend.

Some other observations I'd like to throw into the mix for consideration. (Pedro, you are correct in that these will be weighed from a U.S. viewpoint. It would be interesting to see if these observations are mirrored elsewhere):

For a localized ISP operating in one geographic region, which can economically touch one peering facility, but not necessarily a second, it is difficult to quickly transition from a transit to a near settlement free peering model.

a) One factor is cost itself. The cost of actual leased circuits are now a substantial part of the equation. In many cases, it now represents 25% or more of the purchased bandwidth. Peering tends to mean an increase in leased circuits, there purchased transit can minimize the number of circuits needed. It is unclear what the current turmoil in the telecom industry will do to peering requirements. The large globals such as UUNET, ATT, C&W, and Sprint still represent a large amount of Internet traffic. Should any of these tighten their peering requirements further, it will further impede the growth of an ISP trying to break free of the tran-

sit model.

3. There are signs that purchased bandwidth costs have at least hit bottom. I'll argue also that the purchased wholesale model has completely broken. Add in what is obvious moves of the peering facilities to become financially stable themselves (i.e. increase cost of presence). Does that mean the models we build now are as good as it gets? Since no ISP can keep changing its connection architecture every six months, what variables can be modeled to attempt an accurate portrayal of the future (if any)?

The above observations will form a good basis for examining whether the arguments on the peering-vs.-transit issues have shifted just in the past few months, and whether we need to dispose of preconceived notions and start all over. I expect we'll all cling to our current views, even in the face of certain facts that do not support them. Mostly because we all like to think we see a pattern in the industry, and think we know where things are heading. I'm looking forward to the philosophical end of this discussion, but also hope to see a lot of tempering with current real life experience. Lets get started!

Dileep Argawal

I am the Founder and CEO of WorldLink Communications, the largest ISP in Nepal. I started WorldLink in the Summer of 1995 as a college Junior. After completing my degree in 1996, I returned to Nepal and managed the company till Jan 2000. I was enticed to become a part of the dotcom boom in the US and founded a company in association with Sarnoff Corporation in Princeton, NJ. Together, we applied to ICANN for a new TLD. We made it to the final list of top 8, and were struck down at the last moment due to some minor technicalities. After the illusion of the dotcom boom had faded, I returned to Nepal and have been here since.

Until a few years back, there were only 3-4 ISPs in Nepal. There was a feeling of mistrust and ISPs did not co-operate. However, recently the ISPs have joined hands on a broad range of issues, ranging from dealing with the government to exchanging technical knowledge. At the moment, 4 of the largest ISPs in the capital, Kathmandu are peering over leased lines. This has been effective in reducing transit bandwidth. With the help of Bill and others, we are now working towards setting up a small exchange where all interested parties can connect.

Unfortunately, peering is not as beneficial in Nepal since most of the content accessed is overseas. Besides, transit bandwidth is available only over satellite and is very expensive. So, the thrust is to create useful local content that local Internet users would access. In these discussions, I will bring the perspective of an ISPin a developing country, where the operational parameters are very different from that in the US.

Gordon Cook

The ideas that I want to discuss are typified by the series of packet clearing house NANOG papers by Bill Woodcock and others. The goal that I have in mind is to assess current state of the art thinking on peering and transit. Not just as something technical that ISPs do as a necessary part of being an ISP but rather as a precisely refined economic tool kit that properly employed can make a critical difference in the ISP's ability to function as a profit making business. I suspect there is a body of knowledge that is being developed and is available, in small part, as a few NANOG meeting slide shows but mostly moves by word of mouth.

Note that this emphasis on exchanges for people to interconnect at seems to match the concept that municipal networks will have their own open points of interconnection.

Keith Mitchell in *Carriers' World* July 16 wrote; "It is worth noting that today's low transit costs reflect a current glut of capacity, and not the real underlying costs of transit provision. Even in the face of new optical technologies this remains high. At some point in the future this means the advantages of extensive peering may well outweigh those of cheap transit. Connection to an interconnection platform that allows ISPs to flexibly change the balance between their peering and transit arrangements without major network re-configuration will be critical."

Keith Mitchell

Greetings everyone. For those of you who do not already know me, I've been doing Internet/IP stuff since about 1986, and actively involved in peering/exchange points since we set up the LINX in 1994. Right now and for the past nearly two years I've been founder and CTO at XchangePoint, where we've been working hard to push the envelope of how to run and evolve Internet exchanges. For more details of some of the other Internet co-ordination/regulatory/governance things I've been involved in, see http://www.keithmitchell.co.uk/biography.html .

By the way, the URLof the article in Carrier's World that Gordon references is:

http://www.carriersworld.com/Tmpl/article.asp?CID=8&AID821&SCIDê&TCode=FT&T1)/7/2002

I'm currently quite saddened by the state of the industry - looking back at the past ten years it seems to me the technology has very rarely failed to deliver, and that much of the devastation we see today was caused by over-hyping and exaggeration by the money men of its commercial potential. I've never been shy of commercialization and privatization of infrastructure and associated services, or applying formal structure and co-ordination to voluntary efforts, and indeed see them as a necessary means of getting the technology out of the non-commercial sector to the point where it is useful to most people.

First time I did this was when we set up PIPEX as the UK's first commercial ISP back in 1992. The strange thing is it now seems harder to get buy-in and investment support for good Internet infrastructure ideas now than it was ten years

ago when no-one had heard of it

We set up XchangePoint as a way of harnessing the commercial potential of Internet exchanges as a response to many perceived threats to them, (of which I guess today's basement transit prices are a good example), and also because at the time it was clear we'd been banging up against the scaling limits of running a large exchange as a membership organisation for some time.

The cost of peering vs the cost of transit, and how you can leverage critical mass in both of these to provide the best quality and value for (not just traditional) ISPs have been issues that we have focussed a lot of our attention on at XchangePoint, and I'm keen to see, and be at the edge of the debate about, what the thinking of the other participants is on this topic.

George McLaughlin

I head up Australia's Academic and Research Network, AARNet that connects Australian universities and research organisations to each other and to their equivalent networks and organisations globally See http://aarnet.edu.au and http://arena.internet2.edu

We have equipment at Hawaii and Seattle as well as in all Australian State capital cities. At Seattle we peer with Internets'2 Abilene, CANARIE's CANet, DREN, ESNET, TransPac (to Japan) and TANet (to Taiwan)and through these to the research and education networks of 40 other countries. We have an Indefeasible right to use on Southern Cross Cables Network's trans-Pacific capacity and are looking at taking one on the new Australian-Japan Cable.

There are two different sets of issues:

On the positive side, the various Research and Education Networks throughout the world are working to building a global R&E Network. It goes by various titles, but usually the Global Terabit Research & Education Network (GTREN) - the aim to have a global 10Gbps backbone (including trans-continental) by end 2003 and terabit capacity by 2006. As all parties concerned want this to work, there are no major issues.

Key players are Internet2 (US); CANARIE (Canada); and DANTE (Europe) - but many of the rest of us are involved in various ways. Most countries peer at one or more of the Gigapops of Abilene (US) or GEANT(Europe), eg see for example: http://www.ucaid.edu/abilene/html/peernetworks.html and http://www.ucaid.edu/abilene/html/reachableitn.html AARNet peers at the Pacific Northwest GigaPoP in Seattle. There are now more than 40 countries involved, see the advanced research and engineering network atlas (ARENA) illustrating connectivity and peering http://arena.internet2.edu

On the downside in Australia, Commodity Internet costs are much, much higher than in the US. Telstra dominates the infrastructure and behaves like the Tier ones in the US. There are a number of peering exchanges, but some of the parties that peer here don't take services from Telstra or Optus, and do the rest of their peering via US. This can introduce significant performance issues not only for them, but for unsuspecting others.

We have started to shift our US destined commodity traffic across our own trans-Pacific capacity as that turns out to be a fraction of the cost of taking transit from an Australian carrier. We interconnect to US commodity Internet at Seattle.

By the way - due to the regulatory environment here, AARNet has had to take out a carrier license, which I think makes it the only not-for-profit carrier in Australia (maybe the world) and the only R&E Network in the world that has been required by its national regulator to take out a carrier license.

On a different but related matter, AARNet runs a national QoS enabled VoIP network for its members and into the PSTN (20,000 calls per day). We are exploring VoIP to VoIP interconnection without having to go back through traditional voice carrier SS7.

Philip Smith

I'm Philip Smith, a Consulting Engineer in the Office of the CTO of Cisco Systems. I'm based in Brisbane Australia, with my main activity focus covering the Internet in the whole of the Asia Pacific region. Apart from the usual equipment vendor activities I'm involved in, I'm very much focused on the development of the Internet in AsiaPac. I arrived here almost five years ago, fresh from a five year stint at UUNET in the UK (originally PIPEX - Keith **Mitchell** hired me into PIPEX and was my boss in those early days), and very keen to take a lot of my experiences into an emerging region in the Internet.

To the end of Internet development, I've worked with many fledgling Internet businesses in many countries in the region, helping them spec'ing their networks, putting together the concepts for them to operate their businesses, helping them deal with local politics, and encouraging the concept of "local traffic stays local" in the form of Internet Exchange Points. Also, working with the UNDP/APDIPproject, I helped bring the Internet to Bhutan.

A significant portion of my work is training ISPs in the art of running their networks efficiently, effectively, and with scalability in mind. All of this involves teaching them how Internet relationships are made and fostered, the concept of public competitiveness and private cooperation between ISPs, and very much focusing on how to best deal with maximising revenues within their operational climate. I haven't gone face to face with government regulators (as I strongly believe that force from "white outsiders" is the wrong way to approach domestic problems), but have helped ISPs work with the issues they face.

Aside from these roles, I chair APNIC's Routing Special Interest Group (a working group which discusses routing issues facing the region), APNIC's Internet Exchange Special Interest Group (a newly formed SIG to discuss operational issues facing the IXes in the region), co-chair

APOPS (the Asia Pacific Operators forum), as well as being one of the organisers and Executive Committee members for APRICOT (the AP region's annual Internet Technology and Operations conference - sort of like NANOG, Asian style). And yes, I do the usual round of tutorials at NANOGs, RIPEs, APRICOTs, and any other conference or event which has participants wanting to know more about how BGP works.

[**Editor** asks: why the above doesn't happen more in the US?] **Smith**: Well, from my perspective it doesn't happen in the US because "someone" has decided it isn't needed. Interestingly enough, the workshops I do all over Asia (whether they are under a Cisco badge, or supported by other development activities) are proving extremely popular in the US and elsewhere as a Cisco customer facing effort. I guess what I do is very different from all the certification sort of courses - in a week long workshop, I get down to the nuts and bolts of building an ISP, often including spending one day on building an IXP. Each NANOG I find many many providers who are starving for this sort of content - I'm continuously surprised that my (quite basic) BGPtutorial at NANOG is so popular. After all, the accepted position is that everyone in the US is an expert, and the rest of us are still learning... So no, the US isn't all developed out, but the larger and older providers are generally more experienced, and make more headlines about their complex services. The common thing I meet in Asia is "oh, I want to be like UUNET/Sprint/C&W/AT&T" - pick your provider. But when I see what is happening in this part of the world, the level of expertise is perhaps the same overall.

I publish a weekly "Routing Report" to various operations lists on the state of the Internet routing system. Basically Tony Bates' CIDR Report with a whole lot more added. I don't send it to NANOG, but you can find it at apops@lists.apnic.net, routing-wg@ripe.net, rtma@arin.net and afnog@afnog.org. The main website (kinda broken at the moment) is www.apnic.net/stats/bgp. I have been doing this for almost 4 years now, and it seems to have spawned a little industry of individuals and organisations doing similar work. This is great for the Internet as it focuses minds back on the operational health of our infrastructure.

When I left UUNET I think it was fair to say that we were post-CIDRization - the era from 1994 through to 1996 saw the slowdown in the growth of the routing table to almost linear rates, and with routers getting bigger memory, most people seemed to decide that the Internet would scale for ever. The CIDR Report reported on the CIDRization efforts that people were making, but didn't report on the de-aggregation of the entire address space. Size wasn't the immediate problem, but as Abha and Craig's work highlighted, convergence and stability was more the emerging issue. And more prefixes in the routing table meant that it would take CPUs longer to figure out best paths. Therefore size was still an issue but for another reason. Anyway, I started publishing analysis numbers in late 1998, posting it to mailing lists in early 1999, and expanding it as people requested different views and summaries.

Since then, of the excellent work done, the best example as been by Geoff Huston http://bgp.potaroo.net/, using both his Telstra peerings and the Oregon-IX RouteViews site to actually measure what's going on with the routing table. And many other people are doing similar things on their own BGP tables - Geoff has made his code freely available (I think it still is), so it has stirred interest for others. I should make a list sometime.

Be aware that I'm not saying there is a problem, I just think the Internet community took it's eye off the ball a little in the late 90s - bit like once you build a freeway, you assume that it will last for ever, but in fact you need to watch it, repair cracks, build extra lanes or junctions, etc...

I hope I can contribute something to this discussion (modulo 101 other things competing for my time), mostly experiences from the AP (Asia – Pacific) point of view (and yes, we do things differently here as Bill keeps discovering ;-). As for what I'd like to change to improve the industry, well, the list is too long, so I'll just keep plugging away at the bits that don't despair me too much, and don't involve international politics...

The Asia Pacific point of view is what I was trying to explain at last NANOG - http://www.nanog.org/mtg-0206/transit.html has links. The concept of a neutral interconnect which aims to keep local traffic local is just so alien here. Yes, Bill and I are working on a neutral IX at the moment, and it will be one of the few in Asia. Singapore Open Exchange is the only IX in Singapore, but SingTel's marketing will have told you that STIX is the Exchange in Singapore - in fact it is a transit ISP just like other transit ISPs in Singapore. I've worked with organizations who want to set up IXes to compete with the other IXes - they've missed the point completely; and when I explain what an IX is, they tell me I misunderstand what an Internet Exchange is...! In India we don't have an IX, because VSNLthe monopoly formerly government owned telco (now part of Tata) was decreed to be the official IX for the country. ISPs can only connect to VSNL, not even to each other. Thankfully this is changing now. In Thailand we have two IXes, one private, the other one government, competing with each other - guess which one is popular, much to the annoyance of the other one! (And there has been talk of the government shutting down the private one...), etc etc.

Farooq Hussain

I started out with the NSF International Connections Manager program at Sprint and with SprintLink the company's commercial IP backbone. I was the Principal Investigator for the Sprint NAP and moved shortly after the NSFNET transition from Sprint to MCI joining the team directed by Vint Cerf. I've been involved with peering policy since the planning for the NSFNET transition both with Sprint and MCI. I left MCI just prior to the completion of the merger with WorldCom having worked on both the merger

plan with BT and subsequently WorldCom for the Internet components including a brief time with Concert when MCI and BT were planning an integrated global IP backbone. I was with AGIS for a little over a year helping to establish a business relationship with Telia of Sweden who subsequently bought AGIS out of bankruptcy. The year before the demise of AGIS I had moved to become president of Mediagate Inc. an Israeli-American unified communications platform vendor. Currently, I'm a partner in a research and consulting firm Network Conceptions together with Phil Jacobson [also an ex-MCIer]. We're focused on providing independent research to institutions, vendors and carriers internationally. As an advisor to carriers I'm actively involved with issues impacting peering policy internationally.

Stephen Stuart

I'm Stephen Stuart; currently I'm VP of Engineering at PAIX, the exchange point business that Brian Reid, Paul Vixie, and I founded within Digital Equipment Corporation's Network Systems Laboratory in 1996 or so. Before that I had several years with DEC doing a wide variety of things: CASE software, trading floor software, digital ad insertion systems for cable TV, the NASD backbone network. After founding PAIX and getting it on its feet, Paul and I built a transit-resale/backbone-outsource business that was later sold to MFN, which had bought AboveNet which had bought PAIX from Compaq which had bought DEC (it all parses out, trust me) - and so I wound up being in the same corporate entity as PAIX again. At MFN, I was VP of Research (or variations on that title). Most of my time at MFN was spent re-architecting and growing the IP network, deploying the computing infrastructure to support that growth, and trying to develop new products to leverage the newly-grown network; on the PAIX front, new sites were opened, and metro-area networks for peering deployed in the SF bay area and the Vienna, VA area. Most recently, MFN has filed for chapter 11 bankruptcy protection and is in the process of selling PAIX; at about that time, my dual role serving both MFN and PAIX resolved itself into a full-time position at PAIX.

In my copious free time, I build and operate the ISC network (which contains f.root-servers.net and is the distribution point for NetBSD, OpenBSD, OpenDarwin, the Linux kernel, and ISC software like BIND), advise the people who run the Palo Alto Unified School District network, and agitate for the Palo Alto's fiber-to-the-home project people to get my home on their map.

My primary interest is to chart a course for PAIX, both in terms of getting it safely through the process of being sold to a new parent company, and to maintain its standing and reputation in the interconnect industry.

Mike Hughes

I'm Mike Hughes, and I currently hold the position of being technical lead for the LINX in London (http://www.linx.net/). My official job title is Network Architect, but in reality, my remit is fairly wide, covering representation of the LINX at places such as NANOG and RIPE meetings, plus public speaking and PR duties.

I'm also a co-chair of the European Internet Exchange Working Group (EIX-WG) at the RIPE meetings (http://www.ripe.net/).

The LINX has the enviable (or unenviable, depending on how you look at it!) task of being one of the largest exchanges in Europe. Our current peak traffic is around 16.5 Gigabits per second, and to cope with this, we became the first exchange to operate a standards-based 10 Gigabit Ethernet network, back in March/April this year.

LINX is a membership owned exchange, with 125 participating networks, and employs a staff of 22 people. As such, we have managed to remain "buffered" from some of the immediate pressures and uncertainty resulting from the downturn in the market. We don't get in the analysts' cross-hairs this way.

(Just for the record, I hope that PAIX finds a steady and secure backer or owner to take it into the future, and soon.)

This next bit is aimed at those who are unfamiliar with the European scene. The LINX differs from many US exchanges in that we don't own or manage any co-location operations ourselves - we're just providing the interconnect. Instead, we've installed our switches into a number of different co-location centers in London. From experience, this seems to be a concept some US-based operators have had trouble with at one time or other, despite the fact that this is the "norm" in Europe, and has been for a long time.

Some of these locations, such as Telehouse Europe, could be considered successful transit exchanges in their own right, in the sense that all the major players come together under one roof, run cross-connects, and buy and sell transit (as well as connect to the LINX). This is similar to the sort of environment Equinix are trying to build in the US today - yet Telehouse is not a new phenomenon. It's been this way in London since the mid 90's.

In the current market conditions, we are starting to find a lot more people are using the LINX switched network to lash up rapid solutions to an upstream going out of business - just getting more routes across the LINX from their new provider - although these arrangements generally migrate to a private circuit. The strength of a shared fabric was also demonstrated in the aftermath of September 11th. A number of LINX members who were unaffected by the attacks were able to offer emergency transit to those who were by using the LINX fabric.

So, to the state of the industry. Well, I find myself in agreement with much of what others have already said. A lot of clueless people came along with some unsound business ideas. Most of them listened to vague and wooly consultants. PwC, Anderson/Accenture, DnB, BAH, etc. They all have a lot to answer for, from where I'm standing.

I often get contacted, sometimes cornered, by consultants who just don't get it, who think that buying transit from a "few good Tier 1s" is always better than peering. Evidently, they had just eaten an InterNAP white paper :-). I hope that I managed to steer them in a better direction.

During the boom, most companies "bloated" - hired too many staff, and built out rapidly (everybody had to be global or "pan-European", right *now*) rather than organically, hoping that supply would create demand. They took a calculated risk, but got their math wrong, and were careless with the cash to boot. I read a statement saying the "cost of transit is low" - I think that meant to read that the "market price of transit is low" - lower than what it actually cost to provide, in some cases. I think the payback for this is happening right now, with more to come in the post.

So, what of the future? I find myself nodding in Roxane's direction. While the large incumbent telcos are inescapable, they are mostly too elephantine to be flexible. Even their good techies are bound in corporate red tape. It's going to be up to the more versatile, lighter weight, operations to deliver on this one. Let's face it, they stand more chance of giving a damn about the industry and their customers, because they are less likely to worrying about how to buy that new Porsche/BMW/Merc/Learjet.

Avi Freedman

I am the founder of Netaxs, the first ISP in Philadelphia. I sold Netaxs to FastNet in 2002. My weekday job is VP and Chief Network Scientist" at Akamai. I am also working with FastNet as a 'weekend job'. I participate here as an individual.

John Brown

I currently run an infrastructure consulting shop called Chagres Technologies, Inc. I have another company that is finalizing its ICANN Registrar approvals and will sell string labels to service providers only. Prior to that I ran a rural ISP for about 5 years. Learned bunches from that experience and now try to share that knowledge with others.

Current projects include: Building an IX for New Mexico. This has been a seven year "dream" project. 80 percent of traffic with source and destination in NM leaves the state and goes to EDT, PDT, or ORD. We need to fix that. Current plans include dragging a DS3 to Equinix LAX and interconnecting with various folks there. Then backhauling the peering traffic to NM and its local providers. Sort of a "coop" service. We also have a potential to link via 25 Broadway and then borrow some connectivity over to LINX. I expect this to be operational sometime late in the year.

I'm also one of three peering coordinators for AS 20144, which currently includes IANA, L.Root-Server and ICANN networks. This will be split up shortly and AS20144 will be dedicated to L.Root-Server with ICANN getting its own AS. At least that's the technical plan. We shall see how the Layer 8 of it all plays out. I don't do the layer 8 [political] stuff.

Lets see, what else. Hosted NANOG twice (lots of fun, lots of work). Hosted an ICANN meeting once. Speaker at various other events and now becoming active (passive at the moment) within several IETF WG's I'm on the Advisory Council for ARIN and attend RIPE / APNIC meetings about once a year.

Jere Retzer

I am one of the founders of the Northwest Access Exchange at www.nwax.net in Portland, Oregon. Our university teamed with Portland State University to establish a metro/regional exchange in order to improve the quality of local access. We became operational in December 2001 and have been moderately successful so far with 15 connected networks and a reach throughout Oregon and Washington. We are currently evaluating our business model in order to decide where do we go from here Å\ services, fees, locations, etc. We were not thinking of 'transit' and 'peering exchanges' when we started but have always had it in mind that members would access transit via the exchange and that at some point QOS would become important and for that reason created a fairly high-end environment.

Ralph Doncaster

I run a small ISP (Istop.com) in Canada. We mainly provide ADSL services in Ontario & Quebec. After getting fed up paying about $400/mbit for transit in Toronto (151 Front) I bought an OC3 (less than $2K/mth) to 60 Hudson and threw about $10K of Cisco catalyst gear on each end so I can run FE. I also was planning to get a connection from 60 Hudson to 25 Broadway for NYIIX peering. With the cost of transit now so low, I'm waiting for Telehouse to drop their pricing below $1K/mth for rackspace + 100M to the peering switch. I'm already peering at TorIX (www.torix.net) and OttIX (www.ottix.net).

My perspective is trying to do things as cheap as possible. Up here in the GWN retail pricing for a 1M ADSLconnection is C$30-35/mth (US$19-22), with the DSLAM port lease from the ILEC costing more than half that. Every dollar I save on transit goes into MY pocket. I don't know of any network architects that are paid based on how much money they save their employer, so I think that gives me a significantly different perspective than people who are just an employee.

James Spenceley

I've been involved with the ISP industry in Australia since 95' in various capacities and sadly having had the displeasure of running my own business. Since those days, I found it much better to play with bigger toys paid for by other people.

In mid 99'I joined an Australian startup IP Backbone - COMindico (then IPtel). My early involvement was 'educating' a

bunch on incumbent telco executives (mental note - don't try that again). My main push within the company was to keep the network simple, sadly I missed the message to keep the products simple also, so we are in the interesting position of having a stable working simple network with gradually increasing complex products.

Aside from that, I manage the design of the network (covers most of the country (71 POPs), the purchase and selection of BW, business case for purchasing vast quantities of IRU capacity, build out of our US network and manage both domestic and US peering. We now account for about ~15% of the country's Internet, and have actually used much of that vast quantity of IRU's. We peer quite heavily in the USA (about 60-65% of our traffic peer'd). I forget who quoted in on NANOG, but the first 1/2 is real simple compared to the next half, so true.

Outside of the technical aspects of networks, peering and the industry, I find the general trend in the industry regarding pricing, consolidation and business models (or what appears to be more accurate be 'business hopes and prayers') fascinating. This industry is more like a slowly evolving soap opera everyday. Will Genuity be killed in that car crash? Will Worldcom ever awake from her coma and if not who will get custody of UUnet? Is C&Wpregnant, if so what will she do with the 6 adopted orphans?

Joe Klein

I started working part time for a local alternative paper as computer consultant, layout guy and business advisor. One of the music writers had stumbled upon Usenet and got me hooked up. I became a Usenet News junky. The lead me to the Internet.

By 1992 I was the system administrator for Marquette University's department of Math, Statistics and Computer Science and fully on the Internet. Marquette provided a ton of UUCP feeds local businesses and in May of 1993 I ended up capitalizing on those relationships by becoming the first employee at Alpha.net Wisconsin's first commercial Internet provider. Alpha.net was one of the first regional commercial network at AADS. While at Alpha.net I negotiated the MLPA used at AADS, thanks to help from Jessica Yu and Mark Knopper. Alpha.net was later eaten by PSI.

1994 I was the first employee at NAP.NET (more promised stock) we built a national backbone (all be it a flimsy one) in about six months. I negotiated peering, did IPallocation ... you name it, I did it. I quite NAP.NET a few months before the sale to Genuity over items promised to me including trips to NANOG meetings. After NAP.NET I consulted for a bunch of Midwest ISPs after that Pushed a few business plans under my consulting firm "Titania Corporation" but never got a good VC.

In 1999 I worked as Network Architect at Cumulus and managed to tie together most of the 200+ radio stations in the Cumulus network together before internal politics resulted in the whole thing being dismantled. By 2000 I was the second ISP/network guy hired at Cogent - Director of IP Engineering. I was involved in the transit negotiation and peering. I got Cogent it's ASN, /20 address block (a justification more than a foot thick), turned up the first BGP session. [Next I] joined my old NAP.NET compatriot Phil Crenshaw, then VP of Network Engineering at Adelphia. He also recruited Chris Icide. Adelphia turned up a new OC48c (all Juniper) Backbone in August of last year and started peering in February 2002.

I am Adelphia's peering coordinator and also am attached as the manager of the systems group. The systems group builds the network measurement systems so we have netflow on all peer, transit, and access points. We found a ton of peer to peer traffic and I have been quite open in advocating strong peering relationships between MSOs (Cable Systems). We do about 4.3 Gig at peak inbound from peers and transit. Our ratio of in to out is 2:1. We are somewhat open on sharing our flow data if you research types want a peek. Always happy to share our thoughts on "what the elephant is" with peers. See http://peering.adelphiacom.net/ for peering info. Chapter 11 adds interesting complexity to my job.

Dan Golding

I started out doing LAN/WAN work for enterprises, before breaking into the ISP world, doing consulting for a BLEC that was offering IPservices - getting them an AS, setting up BGP, moving to provider independent IP space, etc. After this, I went to work for Mindspring, in the Network Engineering group, and stayed on through the Earthlink merger. During this time, I was involved in the relatively modest amount of private peering we were doing.

After Earthlink, I went to work for Net Rail, where I was eventually in charge of the groups doing both IPengineering and peering. After NetRail went bankrupt and was acquired by Cogent, I moved to Boston and have been working for a year or so at Sockeye Networks, doing advanced BGP research.

Later this month (September 2002), I'll be starting a new position at AOL/Time Warner, as peering manager, in their architecture group. This will include peering for AOL, Roadrunner, CNN, and all other TW properties. At AOL we do...a lot of traffic, needless to say. Our ratios our very nice, and we have an OC192 international network.

Kurt Erik Lindqvist

I started out with a small BBS system from my original home on the Åland Islands. From the BBS project we decided to grow and established one of two local ISPs on the Islands. While building this we where also working on other Internet related projects and among other things we built the first on-line game where you could win money in real time using the eCash system from Dutch company Digicash. I was later part of merging the two local ISPs into one single company. It was then that I was asked whether I wanted to set-up EUnets operations in Sweden. I became the first engineer and

later Technical Director for EUnet in Sweden. EUnet was then Europe's largest IP provider. After working for the Swedish operations through Qwest buying EUnet International, and the formation of KPNQwest in May 2000, I moved to work for the CTO group where I first participated in building the network supporting the then world largest single hosting environment. In the end of 2000 I ran the network architecture group for KPNQwest as well as served as chairman of the KPNQwest peering group. Following the close of the CTO group in late 2001 I moved to IP Engineering with the responsibility of among other things peering and IP transit services. In that role I worked trough the merger with the Ebone network in 2002 until the bankruptcy of KPNQwest. In June this year I joined Netnod, the national peering point in Sweden, as CEO. Netnod currently operates IXes in Stockholm, Gothenburg, Sundsvall and Malmö. Besides this I have been an active participant in RIPE meetings for several years and for KPNQwest was also monitoring IETFs. I was also one of the founders of NordNOG.

Lauren Nowlin

I am peering coordinator for SBC. I participate from netgeeks.net as a individual only. My opinions are strong and reflect a combination of peering policy at Level 3, ONYX and Carrier1 in addition to SBC. My remarks are descriptive of experience at multiple networks, facing challenges in Asia, Europe and the US, and are not to be taken as explicit policy of SBC Internet Services in any way.

Wouter van Hulten

In the last three and a half years I've helped to upgrade several Internet Exchanges around Europe, in my role with www.interxion.com. We started with building a new colo home for DE-CIX in Frankfurt in 1999, and then developed expansion sites for SFINX3 (Paris.fr), BNIX2 (Brussels.be), VIX2 (Vienna.at), LINX5 (London.uk), SWISSIX (Zurich.ch). Currently we're also building out COPIX (.dk) a 3-colo GigE ring in Copenhagen, NL-ix (.nl) a low-cost exchange in Amsterdam, and MGIX (.nl) an audio/video distribution exchange. In 2000 together with AMS-IX and LINX, I was one of the initiators of www.euro-ix.net in order to promote cooperation between the European exchanges. In 1998, I co-founded www.speedport.com and previously worked for www.research.kpn.com where we built KPN's national IP/e-mail service.

Martin van den Nieuwelaar

Martin lives in New Zealand and produces the Network Intelligence software mentioned on page 2 and pages 17- 18 above.

David Diaz

I'm advising Bellsouth.net (you'll see me on Bill Norton's excel sheet for coordinators) and I was CTO of Netrail.

Use of Looking Glasses and Other Peering - Traffic Related Strategy

Hussain (Aug 3): There's a very interesting part of the interview where BW describes gauging the value of a peer via AS-paths. It is summarized as a doughnut model where Tier 2's find ways to by-pass Tier 1's for transit. Although, Matrix [who as far as I'm aware don't do AS-path measurements BW describes] and CAIDA[who are apparently focused on global traffic flows] are mentioned, there are some other efforts that I believe would be helpful to have his views on at some point. I'm particularly interested to know what the strengths weaknesses are of the efforts whose urls are below:

1)Netconfigs.com http://netconfigs.com/

Woodcock: Netconfigs is essentially a "catalog" web page, listing a lot of individual looking-glasses and tools and so forth. I don't know of any specific methodology that they're advocating over others, though. The little report things they spit out haven't gotten a very good reputation... Everyone I know who's actually requested one has found it to be more inaccurate than accurate, and the analysis it's trying to automate isn't terribly complicated.

2)MTRG http://people.ee.ethz.ch/~oetiker/webtools/mrtg/

Woodcock: MRTG was the initial version of one of the most frequently-used basic building blocks of traffic volume measurement. RRDTool, the more generalized follow-on, took care of a lot of the problems with MRTG which made it less useful to people who were trying to do things other than 95th percentile billing.

"Best peered networks" http://www.fixedorbit.com/stats.htm

Woodcock: No idea who this is... It's basically just repackaging stuff that was long ago done and automated by Philip Smith, Geoff Huston, Hans-Werner, et al.

Hussain: Does Bill feel that his PCH may fulfill a need for Looking Glass Information through the deployment of servers at what seems a very large number of sites.

Woodcock: That's exactly the goal. With support we've had in this past year from Cisco and which we're beginning to get from the APNIC and the European Commission for deployment in their regions, the pace of deployment should increase. We just had a major setback in that Sun just discontinued their only appropriate box, the Netra X1. Please complain to any contacts you have at Sun!

Hussain: The total number of peering exchanges given at 300 is double the estimate at Telegeography - can BW provide some help with the estimate of 300.

Woodcock: It's not an estimate, it's just the simple count of all known exchanges and exchange /24s, and it currently stands at 336. Telegeography just reprints it, filling in any missing street addresses that they can find. I suspect that the reason the list they publish is so much short-

er is a combination of two things: first, it's just the exchanges, not each /24; second, I'm guessing that they don't publish any that they're unable to find current contacts for. One reason I don't know exactly is that Telegeography hasn't been abiding by the terms under which they were supposed to be doing the commercial redistribution of the information, which was that Bill, Antony, and I were each supposed to see a copy of the result each year.

Anyway, the next revision of the list should have quite a bit of good stuff in it, as I'm on the ground in Nairobi and Kampala all this week and next, giving tutorials and doing organizational stuff for African exchange-point folks, and I should hear about a few we don't have any info on right now.

Tudor on August 5: I think that a lot of the guess-work can be taken out of whom to peer with and whom to buy transit from (when you think you need more than two). It is also important to accept the fact that traffic volume patterns and performance change and that today's great peers may not fill your pipes tomorrow. Routing information of potential peers/transit providers (currently available from RIS, PCH, etc.) together with an ISP's Netflow/routing data may be used to perform 'what if' analysis.

Hussain: Let's look at the what-if analysis. How does the cost of peering versus transit work out for ISPs? My expectation would be that this is very region and country specific. Since international transit is an issue that impacts pretty much every country other than the US, the value of aggregating local traffic will be relative to the proportion of traffic that is local (very small for most countries outside of western Europe and the larger Asian and other economies). The costs of peering locally are also impacted by local regulatory and policy considerations. Given that transit costs are likely to be cheap if there is competition, isn't the incentive to seek peering a marginal issue of value primarily for operational and performance efficiencies rather than cost?

Tudor: I agree that some countries have more intra-country traffic than others. But if that is the case it does not make sense to just resell transit, unless you can aggregate sufficient traffic to get a better price from the provider than your customers can by buying directly. But [the what-if analysis] is not enough unless one can peer on demand. On-demand (in the sense described above) exchanges (peering and transit) - if they ever come about -along with 'what-if' information can help ISPs lower per bit costs and reward peers and transit providers for doing a good job.

The initial perception of transit providers I talked to [regarding peering on demand] was that they were going to lose revenue and that they could sell much more lucrative long-term contracts; they also feared further 'commoditization' of transit and increased transparency of their product. Maybe things are different now and some money is better than no money.

Hussain: Peering on-demand: I'm not sure that I understand how peering on-demand works. Could you expand/explain? I know paid peering is in widespread use/acceptance. So is peering on-demand having the opportunity to use multiple paid peering relationships?

Tudor: The basic mechanism is the same, whether you buy transit, paid peering or peer for free. The assumption is that you have a port on an transit/peering exchange and that billing is taken care of. You select from a list of possible transit sellers and fill out the contract details, start date, duration and routing policy. Your peer with a route-server. When the contract takes effect the seller will have changed the routing registry that drives the route-server. Upon receipt of the new seller instructions (via the routing registry) the route-server will give your router default and/or prefixes. As an added security mechanism MAC address filtering and/or VLANs can be used.

Theoretically you should be able to do this with multiple peers in different relationships. Rumor has it that one of the main zebra developers who works for Band-X (surprise!) has been working on this.

Hussain: It looks like there's not much out there to help any ISP determine the value of peering with another network from an independent third party. That seems a bit a of a shame given that there are at least half a dozen efforts under way. I know that the technique and data collected by many if not all commercial networks is considered proprietary and this is thought to have hampered researchers as well as commercial efforts to establish neutral metrics.

Tudor: Metrics such as delay, packet loss etc. can be assessed in a neutral way. 'Value of peering' depends on the traffic mix of the transit buyer with respect to his/her provider and peers, actual and prospective.

Hussain: My sense is that PCH's planned deployment of roughly 150 looking glasses out a possible 336 exchange point locations is to address this issue. Is this correct?

Woodcock: Yes. I described that as my long-term ambition. We're currently deploying at the rate of about one exchange every 3-4 weeks. We'll need to come up with substantial additional funding if we're to increase that rate to the 5+/month which would be necessary to keep up with 50% of the exchanges.

Hussain: If so, it would be helpful to have a more detailed explanation of both the technique to be applied and how the data will be made available to ISPs.

Woodcock: [For the looking glasses,] we're aimed first and foremost at collecting the raw data and making the archive of it accessible to others, rather than providing analyzed results. Long-term, we'll have quite a few "report" functions for simple things like flaps or originated prefixes or whatever, but we're putting more effort into collection than analysis at this point, since there are a lot of people interested in doing the analysis, but the collection has to be

done first to enable them.

Tudor: The major value of the looking glasses (aside from aiding in diagnosing problems, etc.) is to enable an ISPto calculate possible traffic volumes at that IX, should they decide to join it. The ISP would then be in a better position to decide whether to join or not. This is, in a way, a method of sampling before you buy. In the case of tier ones you cannot do that accurately, although you may be able to get some ideas by running your calculations using RouteViews or RIPE-RIS BGP table data.

Woodcock: Or just purchasing the information in the form of the cheapest BGP feed they'll sell you, which is presumably the one that comes over a T1 of transit. Depending on the size of the ISP, this may or may not be feasible as a way of collecting information. Figure $1100 per month per provider whose routes you want. In theory this cost could be divided between all of the users of an analysis service, if providers weren't shy of selling to such a service, and didn't falsify the routes they offered it.

Hussain: Is it necessary to have this data collection effort undertaken by a third party? Given that an ISP should be able to collect and interpret this data by themselves - what would be their motivation to maintain this information as commercial and proprietary?

Woodcock: Typically several "views" of the data are necessary in addition to an analyst's own point-of-view, in order to figure out what the effect of a routing change might be, in advance, which is what we're talking about. A large ISP might be able to maintain external probes for this purpose, but it's easy enough to get without doing so, so I doubt anyone does this. Analysis is something that could be done internally using self-developed tools (which is currently hat happens), externally by an analysis service provider, or internally using a commercially-developed tool.

Tudor: The value of peering with a given ISP is a function of one's customer base, and its destinations predilections. Sometimes you may discover that a certain leaf-AS (non-transiting) terminates so much traffic for your customer base that it is worth peering directly with it if at all possible. An ISP knows who is pulling traffic out of its network, but it is not in their interest to maintain and publicize this information.

Hussain: Is there any value in having a ranking of 'best peered' networks? Namely, does an ISP need an independent third party to turn to in order to determine which are the best peers for it to see either a transit, paid peering or zero-settlement arrangement with? There may also be significant cost and other commercial consideration that would mitigate determination of which peer based on operational optimization alone.

Woodcock: Doesn't hurt, but it also doesn't answer any pressing operational question. In other words, it's an interesting academic question, but doesn't influence the actions of operators. To change their operational decisions, they need to know what will happen specifically from their own point of view.

Tudor: A third party cannot conduct such analysis without Netflow (or the equivalent) information from the ISP who is considering peering or transit. I do not know the legal implications involved when an ISP gives customer information to a commercial entity (as opposed to research).

Woodcock: It's aggregated data, so I think by pretty much any measure, it would be thought to be the information of the ISP, rather than the information of its customers, since the ISP creates the information, and no one of the customers possesses it.

COOK Report: Bill: are you saying that MRTG is really good only for traffic measurement and consequently doesn't do other things that need doing?

Woodcock: Conversation should probably focus on RRDTool, which is where development has been going for the past couple of years... MRTG was an early incarnation.

RRDTool [Round Robin Database – see: http://people.ee.ethz.ch/~oetiker/webtools/rrdtool/] is a very simple building-block which queries any SNMP OID [Object Identifier] and records the result into a database periodically. It's a phenomenon all own its own. It's useful any time what you're trying to measure can be captured as an integer out of an OID. Thus it can be used to count routes, or to count bits through an interface, or to track temperature or memory utilization. But many of these are themselves simple statistical components of more far-reaching analyses.

Exchanges

Mitchell on August 6: [Let me quote and respond to Alex **Tudor**'s statement in his intro discussion above] Alex said: On-demand (in the sense described above) exchanges (peering and transit) - if they ever come about -along with 'what-if' information can help ISPs lower per bit costs and reward peers and transit providers for doing a good job.

I [Keith **Mitchell**] respond that: on-demand peering is actually a much harder objective than on-demand transit, because there is much more negotiation involved in setting up peering relations. You want to demand transit, just put your cash on the table. At least in theory - the bureaucratic hoops required to purchase services from many Telcos these days remind me of the worst pre-deregulation days.

Tudor: I always thought, perhaps naively, that the major difficulty is expressing the business relationship into routing policy and executing the resulting changes without manual changes in the peers'routers'configurations. This holds equally true - I think - for automated transit buying as well as peering.

Stuart: For the networks that I run or have run, one of which had an Actual Peering Contract (written by lawyers), changes almost never resulted in something for the engineers to do. The changes were generally happening in the clauses where "the laws of such-and-

such state or country will apply" were specified.

I would resist contractual obligations that resulted in one-off router configurations if I were not generating revenue from the connection, because varying from the "config." template increases support cost in a way that just doesn't scale.

COOK Report: Stephen, Please would you elaborate here. These above four lines are clearly saying something interesting and I am afraid that I am among the few who can't quite understand what.

Stuart: "Not everything that can be done should be done." While it is possible to configure a network were each BGP peer has a custom policy from a technology point of view (i.e. the router won't crash if you do it that way), it might be extremely inadvisable from an operational point of view.

To expand on that a bit, the more you can build from templates, the better, from an operational perspective. Having every BGP peer fall into one of, say, four or five standard peer configurations is much less an operational burden than having to track tens or hundreds of custom configurations. This is especially important in today's climate, where operations staff sizes are getting cut down to the bone - one of the easiest ways to burn out an ops staff is to increase operational complexity. If you burn out your ops staff, the quality of your network will go down.

Mitchell: [In the sense of on-demand transit] I *think* that at PAIX and Equinix they have tools which allow customer on-line configuration of VLANs for setting up virtual private peering/transit circuits across the switch (also at the SBC/Worldcom ATM NAPs). At XchangePoint we're trying to aim for something similar and currently commit to 24-hour setup of such services.

However, that just covers the underlying switch fabric, and not the commercial relationships between the interacting ISPs. Whether we could ever get to a paradigm where you enter your account number on-line, and use a web tool to purchase transit from a choice of players at an exchange and then have it auto-configured is an interesting proposition - I'm not sure if the Band-X and InterNAPs of this world already support anything like that.

LINX

COOK Report: LINX is a switching fabric connecting exchanges predominantly used for transit?

Hughes: Modulo Keith's comment regarding the "exchanges" - i.e. most being neutral co-lo facilities, selling "housing real estate", with fairly minimal value-add services (most commonly remote hands/facility management) - LINX connects 10 different co-location buildings. The buildings themselves are used for all sorts of tele-housing, ISPs, webhosting, some disaster recovery, enterprise computing, telco nodes/central offices.

COOK Report: If I am buying transit at Telehouse, I can buy a membership in the LINX switching fabric which I would use just for peering?

Hughes: You don't have to be buying transit at Telehouse. You could just be in one of the buildings where our switches are, one of which happens to be Telehouse. Though, a lot of people who do have a presence in Telehouse tend to buy transit there too, because it makes sense to do it - no tail circuit, good choice of readily available carriers.

COOK Report: What is not quite clear to me is whether joining LINX gets one peering with everyone else?

Hughes: Nope. No MLPA here! You negotiate the peerings yourself, bi-laterally. Fortunately, most LINX members have a fairly open peering policy! Many don't even require a peering agreement/contract, a "handshake" will do.

COOK Report: What does the site say? 52% of the global internet available at one whack? Pretty good.

Hughes: We get to this number by comparing the number of routes present in the LINX route collector, which all members must peer with, and announce their peering routes to, with the global routing view we receive from the LINX's own upstreams (this is transit for our own internal networks and back-end services).

Woodcock: Mike, that's 52% of the prefixes by count, then, not 52% of the address space, right? Scott Marcus and I were talking a bit about the difference in the wake of the WorldCom statistics idiocy, and there's a surprisingly large difference. We were going to try to package up the little calculator we wrote so other people could use it, but haven't gotten it done yet.

Hughes: There's a couple of looking glasses here:

http://www.linx.net/tools/index.thtml

Of course, we can't guarantee you will get the same views we do. Some members' peering policies provide for announcing different subsets of routes, based on a regional basis, for example. Others may announce a peer more routes if they pair have a number of different geographically dispersed peering locations - I think Level3 work this way.

COOK Report: So in London there are exchange metropolitan area networks (including, but not limited to LoNAP, LIPEX, UK6X, Band-X. Are these other four similar to LINX and exchange point? What is the market share and business model of these four?

Hughes: They all have varying types of organizational and business models, varying from free, best-effort (LIPEX) to membership organizations such as LINX and LoNAP and commercial operators like BandX (really a Transit broker/reseller) and XchangePoint. The UK6X is an interesting one, as it's operated by BT Exact - this is the R&D arm of BT, and initially started as part of an IPv6 research project - while the production arm of BT (BT Ignite) are participants at LINX.

Take a look at each of their websites for more info. I'm not going to waste band-

width typing anymore up here.

http://www.linx.net/
http://www.lonap.net/
http://www.xchangepoint.net/
http://www.uk6x.com/
http://www.lipex.net/

LINX currently has a dominant market share, however, there is significant overlap between the participants at the various London exchanges, which you will see if you check out the participant lists.

Mitchell: Let me see if I can clarify the European exchange/co-location model a bit better.

Locations that LINX, XchangePoint and other competing exchange operators in London host nodes of their exchange MANs (including, but not limited to LoNAP, LIPEX, UK6X, Band-X), are best described as co-location facilities or data centers (including, but not limited to Telehouse, TeleCity, [Redbus] Interhouse, Global Switch, InterXion, IX Europe.)

In general co-location providers are completely separate companies (commercial but neutral) who only sell space (& remote hands etc) type services - European co-lo providers do *not* in general themselves operate any traffic exchanges for either peering or transit in the sense that e.g. PAIX and Equinix do (though they will certainly run cabling between arbitrary pairs of customers). Rather they rely on operators like LINX and XchangePoint to do this as special-case tenants who bring their networks and customers to the co-lo.

Likewise, LINX and XchangePoint are not in the co-lo business, so in general the European co-lo/exchange provider relationship is quite synergistic. There are also advantages for both parties when one co-lo fills up and there is a need to expand to other sites.

COOK Report: And the advantage of Xchangepoint is that it makes purchase of on demand transit possible??

Hughes: That's the advantage of any shared media where there are people willing to sell transit to those wishing to buy. I don't think XchangePoint is unique in that sense - however, I would be willing to guess that transit exchange traffic constitutes a greater proportion of their load than at the other London exchanges. There's probably a couple of good reasons for this:

XchangePoint actively sells themselves as a transit exchange, the others don't - their primary reason for existence is peering. 2) Many of XchangePoint's customers already peer using LINX and/or LoNAP. Doing transit over LINX as well is a bit too over-dependent on the one piece of infrastructure - Bill **Woodcock** (with Keith's involvement) has done a paper on this.

Mitchell: Making on-demand transit possible is easy and many exchanges achieve this completely by accident. We like to think we go some way beyond this towards facilitating transit sales, making them increasingly easier and more on-demand as our exchanges develop. But at the same time avoiding becoming a principal in the transaction c.f. Band-X/InterNAPtransit re-sale.

COOK Report: But if I want on demand transit just to SUPPLEMENT my peering, then don't I need to belong to both?

Hughes: It seems as though you are making an assumption that you will get all the routes you need over peering. That's probably untrue except for a few providers in the world. I can count them on one hand. Getting your transit and peering over the same infrastructure (and probably the same router interface) is just bad network engineering. You are asking to be bitten on the ass by something. Some redundancy is good, ergo the other exchanges :-).

COOK Report: Globally how unique is LINX? What is preventing something similar from being set up in New York?

Hughes: I don't know of another similar exchange with 10 different locations, or that switches as much traffic on a shared medium. But, at the end of the day, we're yet another Ethernet MAN - though we were probably one of the first.

Woodcock: Seoul still switches more traffic than LINX, right? They're only 4-6 locations, though, depending how you count it. MAE-East never got as big in terms of traffic volume as LINX is today, of course, but it was more than 30 locations at the point at which we tried to cut it over to the MOE-East replacement.

Hughes: I'm not completely sure why something similar doesn't exist in New York. The LINX exists, and was founded, because it's participants want it to. I'm not sure that there has been the same drivers in NY. The major peers have tended to peer across point-to-point telco circuits.

The lack of neutral co-lo in Manhattan is probably a contributory factor too. With the exception of Telehouse at 25 Broadway, the other major co-lo centers, 111 8th and 60 Hudson are broken up into sections, leased from the building landlord and operated by the major carriers.

To run cabling outside of each carriers' segment is expensive, and there may also be "bundling" factors at work - i.e. if you are in a WCOM co-lo suite in 60 Hudson, you have to buy WCOM tail circuits, even to go elsewhere in the 60 Hudson building - yep 0km circuits! The landscape just hasn't been conducive to building a neutral exchange in the same way. But, good luck to the Metro-IX guys in cutting through some of the crap!

Also, the PAIX and NY-IIX had agreed to operate a fiber interconnect, in the same way as PAIX and SIX interconnect in Seattle. As this involves some cross-town fiber (from 111 8th to 25 Broadway), it's taking some time.

Mitchell: I think such a thing is being attempted by Metro-IX, though on a commercial rather than membership basis. I am not completely sure why membership models have not worked for large exchanges in North America, but am sure Bill can offer some insights.

Woodcock: Probably not very good in-

sight. It works fine in some places; Seattle's fairly large, and Los Angeles is medium-sized. There just seems to be a prevailing spirit of "I'd rather buy from a service provider than be part of a consortium" in the U.S. I think it stems partly from there having been too high a ratio of V.C. money to clue for a while, and partly from a reverence for competition and commerce over cooperation in the business climate generally. Same basic problem as lots of U.S. businesses "outsourcing" functions which should be pretty near their core competency. If you look at the way ARIN operates relative to RIPE and APNIC, you'll see it there, too. We have a heck of a time getting ARIN members to participate in policy-making. Everybody just wants to pay the bill and have a low-hassle very commercial relationship with them.

Transit in Wales

COOK Report: The Welsh government just announced a 100 million pound investment in broadband. They likely will be interested in figuring out the economics of getting their own fiber from Wales to the Internet. The LINX is I think the only place that makes any sense for them to go to for peering. But they would have to buy transit from one of the exchanges attached to the LINX or from elsewhere.

Mitchell: Yes. They need to buy transit from somewhere - this could be via one or more long-distance circuits from Wales to the nearest PoP of a national backbone operator, or to locate a router in the PoP of same (non-neutral co-location). However, the smarter thing to do would be to put a PoP in a neutral London co-lo that gives them cabled private interconnect to the various transit providers in the same building, and access to transit and peering exchanges in the same building that can give them multi-lateral connection to a wide range of potential peers and transit providers in that and all the other co-lo buildings these exchanges cover.

Another important part of the model is that the exchange operators do not involve themselves in the bi-lateral commercial arrangements between their customers (though note the IP transit reselling operators like Band-X and InterNAP are different from the pure switch operators here). What this means is that although a particular exchange may state that it is focused on carrying transit versus peering traffic, in practice it is impossible to know for sure (money flow is out-of-band!) and the reality is always a mixture of both.

You have to have transit, as you'll never get to the whole Internet just by peering. Transit always gets you to 100% of the Internet, peering rarely more than 30-35% (the 52% of a fully-peered LINX member is definitely an upper bound here). Transit is therefore mission-critical, while peering is a way of getting your costs down. Peering is less mission-critical, as if you lose your peering you can always fall back (at some cash/performance penalty) to transit. I think it is more true than peering supplements transit than the other way around, which is not to say that losing your peering is not something you have plan to avoid the impact of.

A consequence of this is that there is more value in competing transit exchanges within the same metro than competing peering exchanges. As transit is more critical, it probably makes sense to get it from more than source, where a source is either direct connection to a transit provider, or a connection to a transit exchange. Bill suggests an optimal configuration is two transit and one peering exchange connection in each metro area, but an ISP might decide for example to take just transit at one exchange, and combine peering with transit backup at another.

COOK Report: OK. So once they get to London, they have a LOT of choices. Do you have any thought on how much sense it would make to get their own fiber from Wales to London or more likely their own lightwave?

Hughes: That really depends on how much traffic. Dark fiber from Wales is probably unlikely. Waves are a possibility, and plain old Packet Over SONET (POS) a certainty. Many regional networks take a circuit to London and build a transit/peering PoP in one of the locations there. It's then easy to add new exchanges, private peers and transit providers.

COOK Report: Yes it helps, although as you see, it raises new questions clarification. This peering transit stuff as explained by Bill Woodcock seemed so simple ;-)

Hughes: I think you have to do some simplification to get this stuff down on paper and converted into bits!

Complexities of Peering

COOK Report: But is the reality smashingly complex? Or is Bill really saying in the midst of all this smashing complexity here is one way you can gain some control and comparative simplicity?

Hughes: Going peering gives you a lot more control of how your traffic reaches your network, or goes toward it's destination, if that's what you mean? Exchanges and neutral carrier hotels take this one step further by getting all this under the same roof!

Mitchell: I think like many realities, it only looks complex until you understand the underlying abstractions that are driving it all. Better understanding and communication of these abstractions is what many of the people on this list have (are !-) been grappling with.

Woodcock: I guess I tend to view the routing, the data-path, and the money as three separate but interrelated layers. The parameters do indeed seem fairly simple to me: the goal is always to have at least one route which is associated with a usable forwarding path for any destination IP address, while minimizing cost. Since failure of any system has a probability of 1, most of the work goes into trading off just how many routes you feel like carrying, against the cost of doing so. This all assumes that one is optimizing along logical business principals, not just trying to

keep problems minimally visible while optimizing share price.

So given that, I basically get back to the minimal ISP reference model that I proposed: An ISPis someone who buys transit from two sources, peers as much as possible, and sells transit to customers.

You obviously negotiated the lowest transit price you can, from within the set of prices available from providers who you consider capable of making a commoditized "full routes/usable forwarding path" service actually work, and then you figure out how to minimize the amount of it that you have to buy, while maximizing the number of modulatable bits you have to sell to customers. Which means constantly reevaluating the peering sessions and places you peer in light of your real-world traffic mix.

Which is really the crux of the work in our industry, I think: the algorithm for determining the economic threshold of viability for participation in an exchange. Everything else falls out of that decision as a consequence.

Weller: I'm not entirely comfortable with this model. While it's a fairly accurate description of the growth path most ISPs will take, the path is based more on risk litigation then it is on economics. At some point - probably the next growth phase past this model, risk takes on a different meaning, while economics remain somewhat constant (i.e. stays linear).

We can all agree that peering has the potential to average down the cost of bandwidth. As was mentioned earlier (much earlier - sorry - I'm way behind on the threads here!), to maximize the savings associated with peering, you must peer at the minimal number of physical locations that give you maximum access to a peering fabric. This is more true then ever, as cost of the physical circuits contribute to a larger percentage of the total peering cost.

I know very few regional ISPs that gain anything from peering. They peer at one location, perhaps with a few dozen peers, then smile at how much bandwidth they've offloaded from their transit providers. But if they do the math, they'll see they've not actually reduced costs at all. Only way to get real cost savings is to get significant volume over your peering connections, and that means multiple peering locations, and agreements with the (inter)nationals.

Woodcock: In general, I'm agreement, however it's dangerous to over-generalize here, since the exception is such a desirable case, and we want to understand how to get to the exception. Seoul is the best example of the exception. Local providers can go to one exchange, and offload the vast majority of their traffic. To do that, though, you have to have end-user traffic patterns that strongly favor local content/eyeballs, and that's difficult in undifferentiated first-world English-speaking areas, which is probably what you're thinking of. That's the goal I point every new developing-country IX I work with at, though.

Freedman: I know of some east-coast ISPs who have done this - even recently, but you could probably count them on the fingers of two hands maximum and the percentages are 25-40% of their traffic. And there are still some Europeans who are saving - again - on peering, I think - but only because bandwidth from NYC to LHR (London Heathrow) and AMS (Amsterdam Internet Exchange) is almost free.

Weller: But we also know that peering reduces the risk associated with transit. If I have one transit pipe, and add a second, I have added both bandwidth and redundancy. If I then add a connection to peering partner "X", I've added further redundancy for that particular destination, I've added a small percentage of bandwidth, and I've potentially averaged costs down.

Repeat that same formula, and you've increased redundancy as much as you've decreased costs. At some point, you've minimized your dependence on any single provider to the point where it is acceptable to release your second transit pipe. For the most part you have two paths to every partner network.

Woodcock: I don't know any operators who'd agree with you unless your "for the most part" means somewhere in the 98%-99% range. Not being able to reach more than a couple percent of the Internet means a torrent of customer complaints that can't be blamed on third parties. So what does it take to get 98% of all routes via peering?

Weller: The exact percentage would be dependent on the provider's business plan, but yes, I agree percentage would need to be high for colo, hosting, or business access operations. I also agree, this is not nearly as achievable as it was three or four years ago. Although there still are some smaller companies with near settlement free peering agreements.

Woodcock: I don't think that's [98%] feasible for much of anyone who isn't actually getting 100%. I'd love to see figures from some non-"tier ones" but I'm guessing that few of them are much above 85% or so yet. Though that's undoubtedly climbing relatively quickly, as the noose shrinks around the "tier ones." Not sure whether I like the noose or donut metaphor better.

Freedman: Yes, but there are ways to finesse the move to transit-free with mutual transit to certain peers. It's been done, but I'm not sure if it's been done in the last 1-2 years.

Weller: The few you don't are considered a small enough amount of your traffic not to warrant an extra path. Your largest risk now does not come from the circuits you have, but from the peering exchanges you've chosen.

Transit Pricing

The best thing that AT&T, Sprint, and UUNET could do is drop pricing to the sub-$100 per megabit for low commitments. It breaks the peering economics completely, and most people are comfortable with all their eggs in two big baskets.

Woodcock: If they're losing quite a bit of money at $600/M, what's their incen-

tive to go to $100/M?

Weller: I've not heard the $600 number before. I'd love to see more info on it. Is it hard COGS, or administrative costs on their part? This is a somewhat important statement, as I don't think the peering exchanges are competitors to each other, as much as they are competitors to the major transit providers. If the $600 statement is accurate, then its almost a certainty we'll see a *significant* upward trend in transit pricing over the next twelve months. Any business provider that has multi-year customer contracts out of line with their transit agreements will want to make some sure they are working on a peering architecture significant enough to average down costs to subsidize those customers.

Woodcock: Get a T1/E1 quote from any "tier one".

Weller: Gawd, I wish! Its hard to be competitive even at half that number in the mid-Atlantic area. But I made an assumtion that since we already established that the sample provider has two transit pipes and is branching into peering, then the transit bandwidth from at least one of the providers is probably DS3 minimum. At that rate, tier-1s are sitting around $150-200/M. Again, mid-Atlantic area. Other parts of US, and international, YMMV. Sorry for the confusion.

Freedman: [incentive to go to $100/M?] Stupidity. Who knows, but there are still major IP carriers coming in at or near that price. Maybe they think they are bursting aggregates or something. Hard to tell what they think, since in my opinion, their costs are actually LOWER than what their models tell them.

Weller: So the question is how you create a cookbook? Since the ingredients themselves are variables, I'm not sure you can. What you can do is more a flowchart, where your answers to certain questions define the path you take. At the end of each path, you should have eliminated enough variables to then formulate a model. The models themselves should be easy, the real problem is coming up with the proper questions to create the decision tree. Some that come to mind are:

Is brand recognition (i.e. perceived value) important to your primary target customer base?

Is you POP(s) located economically close to multiple transit providers?

Do have the equipment or engineering talent necessary to manage a complex network?

Are you primarily a dialup/residential provider, or do you cater to large businesses/government/healthcare/education/etc.

Do you plan on offering co location, hosting, or managed services.

There are a ton of others. But I think you get the point. It would be an interesting exercise at some point to create a rather comprehensive decision tree and formulate the end models. I think that will tell us exactly where total transit is a better solution, where (near) total peering wins, and what lies in between.

Woodcock: I really don't think it's this fuzzy. I think one can work the numbers, and come out with concrete answers as to whether the cost of a set of routes is greater or less than the benefit it would bring to your network.

Weller: That does not take into account who you choose for transit, and *that* choice will set the economics of the peering decisions. If I'm paying UUNET $250/M, than any peering agreement I make that averages less in cost than $250/M brings down my total costs. If I'm paying Verio $50/M, my choices are different, and probably there are no peering agreements that would average this down enough to make it a priority.

Your algorithm describes (some of) the models that exist at the end of the decision tree. But the tree sets your business path. As an example, consider two extremes. If I'm strictly a business access provider, UUNET transit may be very important to me for QOS or marketing. If I'm strictly a $9.95/month residential dialup company, Verio will suit me fine. How do you set a weighting factor on the perceived value that is the competitive position of the service provider market you've chosen?

The Netflow Methodology

Woodcock: Alex's posting includes pseudo-code for a description of the model we put together for this a couple of years ago. And my assumption is that it's functionally similar to the one that I know Stephen was using for AS6461.

Basically, take your Netflow exports and the set of all ASes which appear in any AS-path associated with each destination prefix. For each bit to a destination prefix, apply to a bucket associated with each AS one point times a variable divided by the distance at which that AS appears from the origin of the path. Tweak the value of the variable high to indicate that you favor aggregation (small number of transit providers/peers, longer distance to each destination) or path length (large number of transit providers/peers, short distance to each destination). Refine by discarding anything in a path that's to the left of a known tier one.

That gives you a comparison of the relative merit of any AS as a potential peer. Redo the calculation with just selected paths, no weighting, and factoring in cost of delivery, and you get actual per-bit delivery costs for all your bits. Modify your routing table with information you glean from a looking glass to simulate your new selected paths after a hypothetical new peering session has been added and re-run, then figure back up to a total cost, to see whether you should actually be doing any specific peering session. That seems pretty concrete to me. Stephen, is that about the same algorithm you use, or do you do something different?

Stuart: Well, the goal was somewhat different than what a typical ISP would

have in mind; often what we were trying to accomplish was to reach a certain traffic ratio with a specific peer, or to reduce traffic overall to/from a specific peer. The structure of the algorithm would seem to be basically the same, though; just with more initial conditions and some slightly different tests.

The focus on economics, though, misses a point of our peering philosophy that is probably not in vogue in today's economic times: peering increases the quality of the network. When we peered with a network, it introduced a direct path between that network and its customers and our network and our customers, such that performance problems could be addressed in a one-on-one fashion as dictated by the peering contract - networks that we reached through peers were subject to performance problems in distant peering/transit connections that affected our customers but where we had no contractual leverage to cause to be fixed. Qualitatively, I can say that difference was important in a couple cases. In economic terms it probably increased the cost of peering (since peering wasn't approached from a strict economic perspective), but it also retained customers.

Weller: Slightly off topic - I see peering analogous to the old farmer's market. As a farmer, I want to minimize my distribution costs, so I go where the customers are. As a customer, I want to minimize my connection costs, so I go where the most farmers have gathered.

Woodcock: Exactly. A lot of folks still don't get that exchanges are where you want to be, as a customer. I think and hope that as people clue into this, it'll revolutionize the transit market. If transit providers never have to worry about local loop, it'll allow them to cover a lot more ground a lot more quickly.

Weller: Fair market economics works beautifully. But there is one other party that has an amazing control over everyone - the owner of the market building - he simply wants to maximize his profits, and he knows it is difficult for both farmers and consumers to organize and move somewhere else. So he has them both walk a tightrope. While I may not like writing big checks to the farmers, I don't necessarily like paying for all kinds of extra fees to the building owner also. At least the farmer gives me an Service Level Agreement

Woodcock: The check and balance to this is that any party can extend the switch fabric of an open exchange into a less expensive building. That's what kept MAE-East going so long.

Freedman: Yep, though LINX is doing fine and doesn't allow that, do they?

Woodcock: LINX has aggressively expanded out of any single building on their own, so there hasn't been any reason for someone else to go in and do it to them.

Freedman: There was a time when they were in one building with multi-gig Ethernet of traffic, but managed to not have people cranking to hook up their own switches, wasn't there (meaning multi-year time)?

Hughes: The LINX was in two locations before we reached one Gig cumulative traffic. However, the current stats system only goes back to April 2000. The old stats were in a system which was pensioned off a long time ago, and no-one thought of grabbing the historical stuff. As I recall, we hit 1Gig/sec sometime around September of 1999, and Telecity Millharbour was already open at the time, and we were in discussions to build at Redbus HX. But that's just a historical technicality.

It's true that part of the reason LINX went out to the extra locations ourselves was to remove the temptation for others to build nasty ad-hoc extensions to our exchange. When someone does something like this, the operator loses control of the exchange, and it affects the overall quality of the exchange.

Who's to say the operator running the edge switch installs equipment of sufficient quality to handle the traffic, and sufficient uplink bandwidth to the switch fabric? Flat Layer 2 networks and that which surrounds them were designed around being operated by and within one administrative domain. The more you move outside of that, the more you risk getting hurt.

LINX also made it "naughty" for people to build such extensions. How was this achieved? A simple rule: One MAC address to be presented to LINX on each port. Anyone presenting more then one MAC address gets cut off. The LINX management didn't impose this on the LINX participants either - they agreed that it was a good idea, by voting in favor of it at a members' meeting. In general, ad-hoc extensions to exchanges are a "bad thing" [tm].

However, a lot of people are talking about connecting to exchanges via L2 pseudowires. It avoids the need to co-locate your own hardware with the exchange. It helps lower the bar to entry for people trying to reach a distant exchange. This, done properly, in a managed fashion, will work OK.

The LINX way of managing this, still being developed, looks like this: (1) Stick to one participant, one port. Doing so makes it easier from a traffic management, billing and security point of view. (2) Test the pseudowire implementation to make sure it doesn't leak junk frames at your exchange – in other words that you only see what the participant injects at the other end of the pseudowire. (3) Treat the distant participant the same a co-located one: Apply the same technical rules - one MAC address, no STP, no other L2 junk, etc. Charge the same port fees. We're still working on this, and will probably run with some trial participants soon.

I know there could be temptation to buy a pseudowire to a big exchange, put a switch on the end of it, and resell the exchange facilities. Right now, LINX are saying this is a no-go for ad-hoc extensions. Speaking personally, it could be interesting in a managed fashion - in a "Peering by PAIX"-style.

Freedman: And LINX always ran a much more stable layer 2 than the MAEs

- by 3 or 4 orders of magnitude. Of course, one issue with the MAE - the reason we made the "Metropolitan Avi Exchange" and connected fddi/ethernet concentrators and a gigaswitch, is because MFS stopped selling MAE ports because UUNET told them to.

Global Peering and Transit Issues

Hussain: For my part, I believe that peering and transit have evolved into issues primarily impacting liberalization and development in regions and countries outside the larger economies in Asia, Europe and North America.

The gap between the large global IPtransit networks [who directly interconnect to each other and have moved off public peering points] and the rest of the Internet service provider industry is wide enough to make the requirement hurdles commercially unrealistic for everyone else. The fact that zero-settlement peering is essentially off the table and that the choice is between cheap transit and cheaper 'paid peering' doesn't necessarily spell disaster so long as service providers can obtain access to transit networks in manner that allows them to operate competitively and profitably.

At this time I believe that it is an open question as to whether the telecommunications industry can be relied upon to self-regulate itself or whether protection is needed to allow competing smaller networks access to the higher level global transit infrastructure. I realize the irony that I believe we're facing is something quite similar to traditional telephony access to the local loop except in the IP/data network side the access requirement is for the long haul access. Also I realize that on the long haul transit side the industry doesn't have the advantage of half a century of public investment to protect and leverage - rather half a decade of squandered private investment without a clear picture as to how it all shakes out.

Woodcock: I certainly haven't heard much about people buying incomplete sets of transit routes lately... Am I out of touch on that one? Are there still people who think there's some reason to do that?

Weller: Some providers like Level3 and Cogent will offer this - basically equating it to "off-net" and "on-net", where the definition is really "who I have to pay transit to" and "who I have settlement free peering". It's not a bad idea. You have one good global provider - if there is one left :). You then have a few partial providers who together form your second path set. Since each of the providers will offer full routes for slightly more, you've also mitigated risks by having a third (or more) backup pipe should primaries fail for either financial or performance reasons. The problem is that most of the Level3s and Cogents of the world have the same basic peering arrangements, and they don't consider (or can't consider) their connections to the Tier 1s to be "on-net". So you end up with some big holes.

Woodcock's Synthetic Path Analysis

Smith (August 25): Well, this is an interesting analysis, and sitting out here in AsiaPac, I don't really see much of this sort of thing actually happening or being that useful. Yes, I get a lot of questions from ISPs here about who to connect to in the US - and the solution always comes down to money. Who will give me transit access to the Internet, the biggest capacity, for the lowest money?

The result of this is that I'm usually surprised at which providers are chosen as transits - quite often they are further down the food chain than people would expect. Which makes the configurations these providers are trying to do somewhat complicated, and often with less than predictable results.

Some of the bigger providers who are getting their regional acts together (Reach, SingTel, etc) are much more careful about how they choose their upstreams - and they are very definitely doing the sort of analysis about who would give them the best transit, least cost, fewest hops, best service quality, &c. I guess it's down more to Internet maturity.

And we see other providers doing what we did in Europe back in 1994 - landing in the US with our own equipment, establishing a PoP or two there, and choosing our peering partners/buying transit as appropriate. I've been encouraging many ISPs to do this, with great success - once they are sold on the benefits of controlling their own international capacity, and once they appreciate the choice of transit and peering partners they can have, they move ahead at great speed and really wonder how they managed to function before. (The down side with this push is that it doesn't help develop a regional AP network any - it doesn't hinder it, but most people are still focused on "Western content", or their own cultural content sitting in the "West".)

By the way, my untested theory is that because the US West coast is still mostly Asia's peering/transit point, service providers care about getting there at lowest cost because they have a bundle of local traffic to shift. Once they develop local interconnectivity, whether through IXPs, or a better meshed transit provider network, I think their international needs will probably be reduced somewhat - and that might allow them to become a little more picky on how they connect and to whom.

In summary, yes, I find this useful analysis work, but what might also be useful is looking outside the US peering mesh and investigating the reasons why Europeans, Africans, Latin Americans and Asians connect to the ISPs they do.

Freedman: The main comment I have on this piece is that a lot of this data is relevant to the local network where the Netflow is sourced, but shouldn't be generalized as the percentage of the Internet that various networks 'own'.

COOK Report: Can you be more specific as to what part or parts of the interview you are referring to?

Freedman: The specific tables. Genuity is rarely that large as a transit AS in the data we see, as one example. Otherwise, taking Netflow and figuring out which of the transit ASs will peer with you, and costs to get to those ASs, is the standard way people look at peering costs.

Some networks have/had more aggressive peering policies based on projected growth or marketing strategies, and some networks can't easily get Netflow stats so they wait for networks to contact THEM and supply Netflow stats.

COOK Report: Why can't some networks easily get Netflow stats?

Freedman: Because it can destabilize routers. And the OC12+ routers can't do per-packet - just sampled - netstats, which decreases confidence in management in making a measurement infrastructure sometimes. Most importantly, it takes a person who understands sysadmin, basic routing, and basic programming to tie it all together (get netflow exported, aggregated, reporting up, etc), which is a scarce commodity nowadays in the EMPLOYED labor pool(and always was somewhat scarce).

COOK Report: are you willing to apply any of these insights to FastNet? We heard from Phill **Weller** that it really doesn't peer...that it just buys transit. Has this always been the case? Is it a function of current cheap transit costs?

Freedman; I'm going to let Phil speak for FastNet's future direction, but there is still significant peering going on in the FastNet network due to the network merges with Applied Theory and FastNet. (This is visible from BGP-land to everyone).

COOK Report: Do you or any of you have a feeling on how many do this?? Bill seemed to suggest a dozen out of a few thousand globally.

Freedman: Well, in terms of transit ASs, I think it's in the 20-30 range, maybe 50 if you include Europe. In terms of origin ASs, it's certainly more than 100. Probably over 250. [Editor's Note: Further conversation showed that the discrepancy between Avi's figures and those of Bill **Woodcock** are that Avi is talking about those who do actual path analysis and that Bill is talking about both actual and synthetic path analysis.

COOK Report: How long has Netflow data been available? Before it was available would I have a much more difficult time convincing provider x that I had enough traffic to make peering worth while? Is there evidence that the growing number of exchanges has been driven by the desire to peer? Or has it inadvertently enabled more peering? We used to hear about gradations of peering.... (paid peeing and the like) Has the declining cost of transit essentially rid us of these gradations? Either you pay.... i.e. buy transit? Or you peer — that is to say exchange traffic without payment?

Freedman: Netflow has been available for 4+ years, maybe 6+. Of course, define "available" - often wasn't and still sometimes isn't available on the highest-end routers - stably, at least. I think most of the desire to add XPs has been there for many years.

Paid peering is still done and sold, though since the beginning the network, the side doing the selling has really often sold/viewed it as transit. i.e. The buyer could start sending traffic to 3rd parties. There were ways of doing it before Netflow as well, just a bit uglier and more one-offish, though.

Smith: In my past life I was certainly using Netflow information as a basis for determining who to peer with, and even who to de-peer with. Without looking it up, I'd say since around 1995. Before then, it was harder. Certainly on Cisco boxes, people could use things like interface IPaccounting, but I never really did for peering purposes, and I didn't know many who did either. Can't talk for what the industry really did - I guess it was less well defined then? Dare I say the net wasn't quite so busy, and we all weren't competing quite so aggressively - so peering-for-free decisions weren't quite as cut throat as they are now.

Today Netflow type information is a must have for most ISPs I know and work with. Although gathering complete Netflow statistics at OC-n rates is more than just an interesting exercise, it becomes a real technical challenge, both at the intercept and the collection points.

On September 7, 2002 ***COOK Report***: Bill's Synthetic Path Analysis concept — almost no comment. Why? Painfully obvious? Everyone should just 'do it"? Or is it Not obvious? Someone needs to test it before it is worthwhile spending any time on?

Freedman: I believe it's the same as we've been doing for about three years at Akamai - the same data I based the numbers I gave you two years ago on. We do three columns, by origin, origin + single-homed downstreams, and "potential" - which is always way high - esp. since a given upstream AS can't sink ALL the traffic to/from a given multi-homed customer in most cases.

What we do is:

1) Take more than 10 billion records per day of billing logs

2) Aggregate bytes per IP into bytes into prefix generally announced (we see 100k+ more prefixes than most see due to having 900+ BGP feeds, but we only take prefixes that appear in more than about 25 feeds for this step.)

3) Aggregate bytes per IP into: a) Origin AS - but if a prefix is multi-originated, both get credit for it. An option is to turn it off, and report it separately. So you can get that the Internet is 102-104% of itself this way. b) Origin + Multi-homed: If an AS is single-homed to a provider, also count the single-homed customer's traffic into that of the prefix for this view.

Woodcock: Hang on, so if you've got 100 bytes to some destination in AS 100, with a single-homed AS-path: 200 300 100 — where 300 is also single-homed to 200, do you allocate the points to each of the three ASes? I think this is the same idea as synthetic path analysis, but haven't done more review - but without

using Netflow. We've also done one-offs for network partners using Netflow but not dynamic systems to show it.

Freedman: c) Potential: If an AS is included in more than about 25 feeds for a given prefix (also tunable to eliminate this limiter) in the transit path (not origin but including the next-hop-to-Akamai AS), then include the traffic for that prefix in the counter for the AS.

The main reason we don't talk about these #s is because we do like talking about how puny the nasty old no-peering tier 1s are and how they're all really tiny (6% of the 'net or so at max, and that only for a few of them). Of course, by "potential" traffic it's a good indicator of relative size but the %ages you get are really high compared to the amount of the 'net they can deliver traffic to with quality that's AS-connected to them.

One way we use it - I can log into a portal now at Akamai and show the traffic going to 701 - even when we weren't in 701. Also, 701 + single-homed, and 701 + all downstream prefixes. But basically we started doing it to figure what ASs would be best to deploy in. And if a tier 1 wants to be nasty, tough - I'll deploy in their peers, or customers, or transit providers. Many ways into an AS :)

In short I think this is the same idea as synthetic path analysis, but haven't done more review - but without Netflow.

Woodcock: And you just count it once, not once per path it appears in, right?

Freedman: I think the answer is yes. If I have 300 feeds saying Prefix A for ASPATH 200 300 100, but 200 feeds saying Prefix A for ASPATH 50 300 100, each of 50, 200, 300, and 100 would get Prefix A's bytes once - so no weighting by number of sessions we hear a route on.

Woodcock: Yes, that's the same. Basically saying that if you have the two paths:

200 300 100 50 300 100

then you aggregate them into the unordered set {50 300 100 200} and treat them all equally for the purposes of point-allocation.

Freedman: Yes.

Woodcock: Have you tried weighting by distance from the origin AS?

Freedman: Nope, we haven't. That's an interesting idea. We just use the "potential" numbers as a rough guideline, anyway, since we're trying to get into the origin ASs unless said origin ASs are being really nasty.

Woodcock: The idea (using the above pair of AS-paths again) is that you'd multiply the points allocated to AS 100 by 1/1^X, since it's the origin, the points to AS 300 by 1/2^X since it's a hop away from the origin, and the points to each of AS 200 and AS 50 by 1/3^X since they're another hop further away. Tweak the value of X depending how much you value aggregation over short paths. Or you want to vary that value.

Freedman: Neat. I think for a performance-based view, something like that may be cool (though it's not yet been shown that distance from origin makes a difference, it's certainly common wisdom that it does). For cost-based, it probably matters less, I think.

Martin van den Nieuwelaar: I think the idea is sound. As far as I'm aware, however, there aren't any experiments to confirm how well it works in practice. One has to take the current network routing table and augment that with the routing table provided by the potential peer. Hopefully the resulting routing table you end up with, and the routing that takes place if you did actually peer, are the same or at least very close. Otherwise the simulation is not accurate.

Taking the concept further it's possible to imagine a tool where you import your network traffic pattern, and have a list of selectable peering points. By selecting or deselecting individual peers, you instantly see the changes in transit/peer volumes. One day (*) when bandwidth really does become a commodity it might even be possible to include costs for peering and transit. Then it's possible to let some minimax algorithm run over your data and tell you the most cost effective peering points. By that stage, the prices of peering and transit are probably varying on a day-by-day basis, so all this gets worked out automatically every day, and there's some automated mechanism running behind the scenes to enable/disable peering/transit connections without you having to worry about it.

Of course all this is just a dream. Most people today seem to have difficulty measuring and understanding the Netflow statistics produced by a single router.

(*) This may be a way off. In my experience the price for peering or transit can be based on quite odd things. Like political reasons, where peering with X will give us some advantage so we will/won't charge them etc. Also, translating circuit contracts into equations usable by machines is sometimes tricky. Examples like "break the contract before 12 months and you pay X", "extend contract one year and receive 5% discount". It's too messy to want to think about!

Doncaster: I've used Netflow and have found it not to be useful for technical and practical reasons. On the technical side, you only see best paths. That means you don't get an accurate idea of the benefits of peering with a particular AS - the actual benefits are much more than Netflow analysis would imply in most cases. From a practical view, finding out where your traffic goes is not the problem, connecting with the other network is. Therefore I find it much simpler to connect to exchanges that are convenient and inexpensive, then peer with everyone possible at the exchange.

My traffic levels are currently sub-OC3, but even up to OC192 traffic levels I see the same economic trade-offs. I could buy an OC192 into Buffalo for <$10K/mth. When buying transit with commit levels over 1Gbps all indications are that even the tier-1 networks offer pricing for less than $50/mbps on 1yr contracts. I believe some (like QWest and Level3) are less than $25/mbps at those

kinds of traffic levels.

Economy of Scale

COOK Report: Do any of you have any opinions on whether Bill **Woodcock**'s views on peering and transit and their purported effect on ISP viability have anything to say about what economy of scale in the Internet is all about? Surely Fast Net has a very very different economy of scale and hence viability than SBC?

Freedman: Yep. FastNet can become profitable and SBC can't as easily. Or maybe they both can become profitable.

There are two factors working:

1) Large companies tend to do things less efficiently. Along the curve of wasted overhead/revenue the less than 10million per year companies and the greater than the billions of dollars companies typically do the best - but at current DSLprices, with infrastructure costs and assuming capital is not free, I don't see how SBC can make money (i.e. price can also prevent companies with revenues in the multiple billions from making money).

2) In the peering game, certainly there is more efficiency the more traffic you have, if you can keep the provisioning etc groups sane internally in terms of overhead. Of course, at a certain size the infrastructure is a killer, as 10gb/sec ports are much more expensive than 4 x 2.5gb/sec ports. But in general, 100mbit/sec of transit is cheaper to sink via peering than .5gb/sec is and up the line. It mostly has to do with the efficiency of local loop and interconnect costs.

If you put all of your eggs in one basket and everyone moved to the Equinixes, our Internet robustness would suck (re: resistance to attack), but the extra benefit to scale would be less, though still there.

Quality of Service

COOK Report: Is there enough bandwidth out there, and is inter connectivity now so good that Quality of Services issues are no longer really relevant? Can UUNET any longer say to a large global enterprise you have to be a customer on OUR network in order to have acceptable QoS?

Freedman: 1) Inter-connectivity is not so good that QoS doesn't matter. Of course, no one can offer QoS (see 2 below)

2) UUNET's argument has always been specious, since access pipes to UUNET's customers are sources of congestion as well, and since UUNET has been pretty good in the last year or two about keeping open peering paths to those they peer with.

COOK Report: What kind of applications are no longer QoS dependent in the sense that they can be counted on to just 'work'?

Freedman: let me put it this way. Packet loss sucks. It is to be avoided. No network avoids it to everywhere at all times. Corollary: All networks suck. Some suck less but most of the routing optimization companies have data to show that most reputable networks all suck some and they all suck about equally.

COOK Report: So there is enough bandwidth out there to do most any reasonable thing? VoIP? Streaming audio or video? I am talking about what someone in an enterprise might want to do — not someone in high energy physics. The problem is that changing network conditions.... burstiness, etc will lead every now and then to slow response or even packet loss? Is this really a sore point with anyone? Do enterprises avoid it by tunneling connection oriented VPNs through the public Internet?

Freedman: No, that's not what I said :)

Quantity there sometimes is, certainly much more so than a few years ago. Quality there generally is, but for any given endpoint to endpoint it won't be there long-term unless both endpoints are on one network and you have QoS at the endpoints. Use X application from your hotel room to a university and all bets are off - for the next 3+ years I bet.

The problem is that most applications and users will hit congestion at times, and the end to end QoS problem is really insoluble today at the IP level.

COOK Report: Given the alternatives, don't the rest of us pretty well just have to live with it? Does anyone here see a pressing need to reorient the transit, peering and exchange point models that are being discussed JUST to try to solve QoS?

Freedman: Well, some people would like to see that but it won't happen. The most that's been happening is that 4 or 5 core networks looking inter-provider QoS, and even then that's well under 1% of the significant inter-AS links.

COOK Report: And Avi just said that short of going to IP over PVCs (connection oriented) QoS isn't solvable in a packet switched network? Is this what you are saying Avi?

Freedman: No, I'm saying that the technology isn't the big deal. It's establishing the business relationships between all of the interesting ASs on the 'net that is the HARDEST problem. I will say, however, that just the billing and settlement for the Internet would probably be 100 billion records per day - not trivial.

AND you need to get everyone filtering forged-source and probably deploying new routers at the edge (w/ no budget) or people can just forge their way to better access. But the technology is doable - the business relationships are the hard part.

COOK Report: You mentioned routing optimization companies of which I dare say you must know a few things about Sockeye. How orthogonal is what these folk do to the peering and transit issues we are discussing? These folk are about load balancing to your directly connected networks? Yes? Is this all they are about?

Freedman: They look at visibility and verification, cost balancing, and performance balancing. All of them have the same message though based on empirical data: Pick any two providers one thinks of as low-end, combine them using some intelligence, and you're better off than

with any one provider, no matter how "premium" (including, UUNET, InterNAP, etc)

Odlyzko on Pricing

On August 21 **Odlyzko**: The question I would like to pose is whether the transit/peering issue matters much or not. It seems that at least in the US, the cost of transit is low enough that it is not quite irrelevant, but should not be critical to the survival of an ISP that can reach decent size. Here are some calculations that support this. The basic assumption is that the ISP can buy transit at something like $150/Mbps per month (and so has enough traffic to fill an OC3, say). [Let's look at two groups of costs.]

1. ISP serving residential broadband customers: The highest figure I have heard for the traffic that such users generate today was for a group of university folks, who were running an average of about 10 Kb/s of traffic (sum of incoming and outgoing traffic) per subscriber over long periods, with relatively low peak-to-average ratio (since most of the traffic was peer-to-peer music swapping). Well, even if we allocate 20 Kb/s for each user, and in each direction, to accommodate growth in usage and higher peak traffic, we can get 50 such residential users into a 1 Mbps slice, so the cost of transit per user will be $3/month. That is simply not that much! (And in practice usage is far lower, and so are average per subscriber costs.)

2. Global computation: Current estimates (by myself, RHK, and others) are that US Internet backbones carry something like 100,000 TB/month, which comes to 300 Gbps (averaged over a full month). Suppose we send that through a transit connection that runs at 30% average utilization (measured over a full month). (This, I am told, is a reasonable approximation of what is achieved in practice, but perhaps somebody on the list will correct me.) That means the transit capacity will have to be 1,000 Gbps, which, at $150/Mbps per month, will cost $150 million per month, or $2 billion per year. (More details to justify this: Think of a far simplified world with just 2 ISPs, call them X & Y, of equal size, with customers of X communicating exclusively with customers of Y, and vice versa. Suppose that X & Y both buy transit from a backbone B, that basically just connects them. The traffic will on average be 150 Gbps in each direction, requiring transit capacity of 500 Gbps for each of X & Y, leading to the cost figures above.) Yet the US backbone Internet connectivity market (excluding modems, etc.) is supposed to have revenues of $15 billion of so per year. Thus most of that revenue is coming from low bandwidth connections that are run at low average utilizations.

Comments?

Note that if you accept this, then we have all sorts of interesting implications for likely structure of the industry, since this says most of the costs are at the extreme edges of the network.

Freedman: I agree that while ISP transit is 150/mo in semi-quality, peering is iffy until you're at the gigabit-ish level - or hundreds of megs with *confidence* you'll be at the gigs level. Semi-quality means good enough that most of your customers would accept it but not the quality you'd like, or that fussy customers would accept.

Roughly the math [for peering] which I see working again and again is:

- If the people are free (almost never true, figure 1 Full Time Equivalent @ 10k/mo) - And the routers are free (OK, true now) - The interconnects/racks will still be 50-70 dollars per megabit loaded costs at this traffic level - IF you already have a inter-regional backbone that traffic can ride - otherwise you have added costs - And you can get 30-40% of your traffic free - in some cases by being only at PAIX/PAO or the east coast usual locations (NYIIX+ MAEE ATM+Equinix)

But you're not going to grow that 40% to 60% of your traffic via settlement-free peering. Not just locally - even nationally without a million per month investment in:

(1) Global infrastructure and peering points to pick up smaller networks (not economical) (2) Going to 3-5 Equinix facilities, participating in the SuperPeering product (forget the exact name) and really having some sort of real long-haul infrastructure. The infrastructure necessary for settlement free peering is decidedly not free.

So is it worth it? Advantages:

1) Geek jollies 2) Marketing 3) Element of control of destiny (and anyone at an XPcan give you POPtransit or complete transit for $ or free in an emergency) 4) Long-term strategy

I think some clueful 1-10 million dollar per year ISPs manage to do peering at the non transit-free level and at least not cost themselves money, maybe save themselves the cost of some unajus and kappamakis (Japanese foods) each month - but big businesses can forget it. Also, international providers may also have strategic advantages from having a US network including reputation back home and the ability to be less beholden to particular transit providers.

And I wouldn't doubt that the right one or two people might again be able to build a transit-free network if they had enough traffic and growth - Cogent has held on to peering a lot more than anyone thought. BUT those one or two people will be too busy getting and keeping customers nowadays and won't have time to politik around for a two year peering odyssey.

Just my aggregated thoughts - lots of this has been already said.

Exchange Economics for Colorado Springs

COOK Report: How might you look at Colorado Springs where these is no peering exchange? Could there be a business case for an ISP opening a peering exchange and aggregating the peer's up-

stream traffic via an OC3 to PAIX Denver where it would connect and provide transit for the nets that were peering in the springs? How do you think about doing something like that?

Stuart: What you are describing is similar to the "ISP Condo" part of the AboveNet business model (that pre-dates the acquisition tangle that caused me to be associated with AS6461; Avi may be willing to add some details). Basically, AboveNet's co-location centers (ISXs) attracted what could be loosely defined as "retail" ISPs (those that would run circuits out to customer locations). AboveNet provided a separate switch for those networks to peer with each other, allowing (encouraging) them to bypass the transit that AboveNet sold them in the ISX when the traffic was exchanged with another AboveNet customer in the same facility.

I remember hearing about what AboveNet was doing at about the same time as Paul Vixie, Brian Reid, and I were starting PAIX, and thinking that the ISX concept and PAIX had some similarities. They both encouraged local traffic to stay local, and bypass some expensive thing - in AboveNet's case, traffic avoided the upstream transit interface where the dollar-meter ran; in PAIX's case, traffic avoided the circuit-priced private cross-connect.

COOK Report: Something tells me it would not work but I am not sure why.

Stuart: Technologically, it does work. Economically, it might be a means to an end, but I don't think it would generate enough revenue compared to the costs to be an end in itself.

COOK Report: Could the peering exchange in the springs use its circuit from PAIX Denver to the springs to backhaul Denver ISPs that were looking for a cheap place to peer? However, the Denver ISP would have to have a router at PAIX to climb aboard the circuit to the springs, so even if the ISP put a box in the Springs it wouldn't save anything would it?

The Springs would work only if there were enough ISPs in that area to connect locally as spokes into the hub and connecting locally were cheaper for them to do rather than connecting 65 miles further north in Denver?

Stuart: I know of one case of a telecom provider doing long-haul L2 to extend PAIX ports to customer connections quite far away.

Freedman: There are a few ways this could work. Most likely, the ISP opening up would just sell transit, or sell transit to peers at PAIX. The latter is harder, and less likely in general, though things like that do happen. Generally when someone sells access to peering it is done in-building or in-city.

Or, the ISP could sell Layer 2 to PAIX, and then work with PAIX to sell a package of Layer 2, peering port (or not), and 4U of space to put a router and a switch (the switch for PNIs).

Menard: I was wondering the basis upon which peering was executed at most exchange points with regards to measuring traffic. If it is really the concept of equals peer freely, then a smaller ISP should have all the motivation in the world to recruit subscribers who are bandwidth heavy on the upstream so as to return as much traffic as it gets? Does this make sense (i.e. end-users upload big files into other networks) which gets an ISP to peer for free?

Woodcock: Yes, this is common practice. ISPs will give away unidirectional transit, or limited-route transit, to folks who can use it, in order to make themselves more attractive peers. Occasionally, ISPs will generate artificial traffic by simulating end-user load, in order to make themselves look like attractive peers, though this is pretty clearly unethical.

Andrew Odlyzko's Evaluation of the Peering World

Odlyzko: I find the prospects of smaller networks being able to bypass the Tier 1s fascinating. The development of tools, such as those of Bill Woodcock and others (listed by Gordon) is also very interesting, and for several reasons: (i) it should accelerate the evolution that Bill and others are describing, (ii) it offers opportunities for interesting technical research on improving the tools, and (iii) the fact that such tools, or even attempts to do such optimizations manually, have not been used much in the past, confirms the general conclusion I have drawn some years ago, namely that optimization of the usage of physical resources is not a major priority, that dealing with general complexity matters far more.

The last point, (iii), might be worth spending more time on. There are various arguments that support it. For example, given that data traffic is typically asymmetric, and traffic profiles are pretty stable, why doesn't the industry develop photonic and electronic systems that would allow for switching the direction of transmission on a fiber strand a few times per day? Yet I have not seen anyone talk about such systems, although my friends in the photonics area tell me it would not be too hard to do.

Stuart: Would it make the network easier to manage? While the architecture and engineering sides of me think that's a really cool idea, the thought makes the operations side of me cringe. Backbones are hard enough to manage with systems that implement capacity in a static fashion. The problems that could arise with capacity coming and going - even if on a scheduled basis, rather than "on demand" - would (I think) have a big impact on operational expense.

Odlyzko: No, it would not make the network easier to manage, just the opposite. And that is exactly my main point, and the reason for citing this possibility, namely that, in practice, it is the difficulty of managing networks that dominates actual network engineering decisions.

Stuart: Yes.

Odlyzko: (Yes, there is a lot of talk about the wonders of QoS or MPLS, but how

much of this ever gets used?)

Stuart: Well, MPLS is in its third or fourth generation of trying to find a raison d'etre. First was to solve the IProute lookup problem that went away with ASICs. Second was TE tunnels that assumed that provisioning problems would drive providers to take traffic off the IGP-shortest-path. Third is L3-VPNs that are a monument to needless complexity, and I think I've missed one).

Op-ex budgets are *low* right now. Technologies that promise to deliver op-ex savings generally do not behave that way when they're in "early adopter" stage - and that's going to be a difficult hurdle to clear in today's economy.

Odlyzko: Are op-ex budgets low? That is the opposite of my impression, that op-ex is an increasing fraction of total expense.

Stuart: Yes, op-ex budgets are low. It would not surprise me to find your assertion true as well, since cap-ex and total budgets are plunging as well. Op-ex could both be dramatically down relative to such spending two years and still be a larger fraction of total spending due to cap-ex money having dried up and blown away.

Odlyzko: Another point is the low utilization of backbones. Gordon in his Introduction mentions an estimate of 15 to 20%. My guess is that it is probably lower, more like 10 to 15% (and conceivably far lower, because we have all those new players that have built out their networks in the anticipation of a huge growth of traffic that has not materialized). As an example, consider AboveNet, with traffic data available at <http://west-boot.mfnx.net/traffic>. During the week ending Aug. 24, the average weekly utilization of their long-haul links was 11.5%. I have a variety of other arguments for (iii) in my papers.

Woodcock: Does this seem like it's come down from numbers in the 50%-70% rangethree-four years ago?

Odlyzko: I don't believe the AboveNet average weekly utilization was ever in the 50%-70% range. (In fact, I have not seen any network that has operated in that range for any length of time. In one direction, yes, but not when both directions are considered. If somebody has some counter examples, I would be interested in hearing of them.)

Here are some weekly average utilization figures for AboveNet (March 1999 is the earliest data set I have):

date	average utilization
1999.03.31	18.2%
1999.07.02	16.4
2000.02.26	28.9
2000.07.02	11.6
2000.11.26	10.6
2001.04.14	10.1
2001.08.24	6.5
2001.11.19	8.4
2002.01.27	9.8
2002.06.07	10.6 - 11.5

Freedman: We certainly had links that had to be used (cross-country links) that were in the 50-70%+ 95th percentile in a given direction.

Odlyzko: Yes, with the crucial provision "in a given direction," which is what I had mentioned before. (Also, 95th percentile tends to inflate traffic measures.) But has anyone ever a large network, not a single link, at anywhere close to such utilization rates?

Freedman: Comparing averages across the whole network isn't that useful from a traffic engineering perspective.

Odlyzko: Absolutely, and I am not presuming to tell people how to engineer or run their networks based on such data. However, they are illuminating when one considers the economics of the industry as a whole.

Economics of Maintaining Capacity

Stuart: Capacity can afford to be low among providers that control their own optical infrastructure, and that if you looked (were able to look) at a provider whose backbone was built out of capacity bought from a telco you'd see the opposite - again, it's a function of the operational expense incurred, and the ability that a provider has to eliminate that expense by eliminating capacity. When we (Paul, Hank, and I) built the version of the "AboveNet" (MFN) backbone you're looking at, we were building a network where the infrastructure right down to the glass in the ground was planned to be dedicated to IP. There is no incentive to retire that capacity just because demand for IP is low (whether industry-wide, or MFN-specific due to Chapter 11, however you see fit to interpret it), because op-ex wouldn't be reduced. Likewise, providers with leased lines (the ones that haven't been driven out of business by providers with their own transmission capacity) have contracts and cancellation fees to consider before they can realize savings by eliminating capacity. It's often better to let the circuit sit relatively idle and pay what is comparatively a small fee every month than to take the cancellation fee all at once - and the contract can always be rejected in Chapter 11 proceedings.

Odlyzko: I have seen network stats for some providers whose backbones are leased, and indeed some of them load those links heavily. It appears to depend on the costs of leased lines, and also on customers. For AboveNet, I have been watching their network utilization for three and a half years, long before bankruptcy seemed likely, and even during the period of rapid expansion, utilizations remained low. (There was a spike at the beginning of 2000, though.)

Stuart: AboveNet/MFN is something of an anomaly in that regard - Dave Rand established, and Paul/Hank/I continued, a policy of "massive over provisioning" (QoS meant "quantity of service"). From an engineering perspective, utilization was kept low due to a combination of "provisioning ahead of demand" and a desire to keep enough headroom in the backbone to accommodate bursts relative to peak. The subsequent market decline provides some desire to rewrite history,

but that is indeed what we thought we were chartered to do at the time.

That last bit, "relative to peak," is important. When times are good, planners can base their policies (like the policy for scheduling circuit upgrades) on peak utilization. When times are bad, the basis for calculation shifts from peak to average utilization; circuit upgrades grind to a halt for a little while because the threshold was changed. The sad thing is that in the typical case, the ops staff was probably also reduced at the time the policy was changed, so all the extra Traffic Engineering that you need to do in order to try to maintain good service at a time when there's less headroom in the network has to be done by fewer people.

Odlyzko: What would be your estimate for average backbone utilizations for various types of ISPs?

Stuart: Don't know.

Does Peering Save Money over Transit?

Odlyzko: The reason for spending this length of time talking of point (iii) is that it is relevant to the main question of this list. I my previous message, I raised the question of just how important the cost reductions from peering are, given the low cost of peering. I have not heard any comments disproving my estimates that once an ISP (in the US, I am not necessarily talking of other parts of the world) aggregates enough traffic to fill an OC3, say, the costs of transit are not all that high.

Weller: Correct. At today's prices, one can show a total transit cost model to be nearly dead even with one that includes peering. When you factor in human resources, it tips scales in transit's favor. Its a shame that this probably won't last more than six to twelve months.

Odlyzko: Why? Can you explain? The precipitous decline in transit prices may slow down, but prices should continue going down. Fiber is now basically a free resource. So it is just a matter of lighting wavelengths and putting on routers, and those things are getting less expensive.

Weller: Its not necessarily a factor of infrastructure. It just seems no one wants to be the price leader these days. I've not seen major Tier 1s break the sub $100 level, even at multiple OC12s, despite the fact that they've been hovering close to it for past six months. As you say, some of the minor Tier 1s are trying to fill empty pipes, but various factors question their stability, and customers want to avoid risk these days.

I do expect T1s to drop some more, as providers adjust for cheaper upstream bandwidths. But I do think Tier 1s will raise prices - or more accurately - not discount them as aggressively, to start regaining lost margins. Or in some cases, to have enough cash flow to emerge from Chapter 11.

Klein:: [as far as Andrew **Odlyzko** calling fiber a free resource] ...until we find that stuff like 10G and very high-density DWDM won't run down some of it, because it's so old, crusty, full of patches and mangy splices. Then we'll be back to civils and guys in the roadways. Hope people have deep pockets by then. (I'm just being Devil's Advocate here, by the way.) But, as fiber is a plentiful resource right now, we can get away using less demanding technology, because the need to "sweat" the fiber is lower.

[. . . lighting wavelengths and putting on routers,] is still the expensive part! Even with surplus equipment being sold off, transmission gear still isn't *that* cheap, especially at the higher ends.

With some financially troubled organizations actually turning whole networks (or just segments) down, could this be constraining transmission resources in some areas? Putting whole fiber rings, PoPs, etc., into limbo until the administrators dispose of the assets? Has anyone done any research into how deep this has to cut until it becomes visible in the market price (not cost, note) of IP transit or clear circuits?

I guess this needs to be looked at from two angles as well, one is hub-to-hub capacity where there is plenty of provider choice. The other would be in the more fiber-remote areas where there is currently a small oligopoly of providers. The withdrawal of a small number of providers from those markets leaves only the RBOC/incumbent PTT and maybe one or two competitive carrier(s). Is that enough to force prices up in those markets (as long as there is maintained demand)? Just a thought.

Woodcock: This actually brings up a different concern of mine that's been bugging me more lately. The fiber that we're all using right now was, for the most part, financed by bankers who were looking at amortizing it with some amount of high-dollar-value voice-minute traffic occupying the extreme bottom end of the bandwidth. All the excess capacity then got sold off at whatever price could be gotten, for Internet use. That excess capacity, in turn, is getting sold off to VoIP providers, who are undercutting the $1 per minute traffic at $0.03 per minute. So what happens when we use up the current capacity, and have to explain the amortization basis for the next round of installs to bankers?

I'm not explaining this very well, but it looks like it might be a relatively large problem five years or so from now.

Spenceley: This mirrors similar concerns I have for the AP region, with the wonderful increase in capacity we have seen in the last the years and the availability of long-term IRU's. The corresponding increase in capacity that has essentially been sitting idle, coupled with the pressures to increase revenue from backers/boards/shareholder et.al. making executives and sales types come to the wonderful conclusion, they have such spare capacity its better to sell it below cost than have it sit idle.

Now in two or more years when that capacity is full and people like SPRINT have sold their STM-1's at below amortized cost (with transit), and the market is completely below any sound price

point, why will anyone purchase new capacity and more to the point will it still be (ROI) worthwhile running the cable, and that is scary place to be. Exactly how do you break that cycle ... after it's happened?

Woodcock: That, or the price of VoIP will stabilize much higher than it is right now.

Spenceley: Will VoIP stabilize, or will it be transit that stabilizes? Does VoIP really utilize that much of the pipes ?

My guess is that transit pricing will mirror RAM, after a period of solid reductions, prices will be up and down like yo-yo, depending on the utilization of the remaining providers links, or number of providers. Finance departments might even start to cost BWbased on daily cost structures rather than 10 years plans. Peering right now might not be the most cost effective option in some cases, but if you are planning for the future, like Bill says I wouldn't want to be totally dependant on the "tier 1" pricing.

Diaz: Then people will go under, or in chapter11 and have the ability to break their contracts. I have heard a lot of providers mention that they plan on raising prices after the turmoil ends. After the sick ones that are killing the market go away, the remaining will raise prices. OR buy those assets for 2 cents on the dollar, changing the cost basis again. It's always a cycle.

Sprint has been looking at the new technology though. There are a lot of places to be concerned. If landlords aren't paid, and the fiber runs through there, or it's a regeneration site and they unplug it, what happens? [For example: MFN fiber?]

Ralph Doncaster: I hear this a lot too, but how realistic is it? With current costs of equipment and dark fiber, you can put a POPin DC and NYC and get a 20 year fiber IRU all for under $8M (possibly even under $5M). That would be with a couple Nortel Optera 1600s or Lucent 800Gs.

Sell OC192 waves @$5K/mth each and your gross is almost $5M/yr on 80 waves. 5 years from now the same fiber will be able to carry 400 OC192 wavelengths or more. Forget about OC3 or OC12 so you don't need ADM gear and your operating expense costs drop significantly. So the long-term pricing trend is only down the way I see it.

Diaz: Well you would be getting hosed. It's already 160 waves forever and there is a new band. Sycamore is cheaper and better, ONI cheaper still (but take your chances). With circuit bandwidth changes on the fly, you can now over subscribe layer1 (not that I would). Redo your numbers and see now.

Matter of fact, you could shoot the new lasers over old fiber and no one would ever see you. So existing players could see u the 15xx band for really cheap and you could still shoot your new stuff down it

Freedman: But there's no market for selling waves. On the other hand you can put POPs in (if you can find space and power in 60 Hudson in New York City) and rent an OC12/48/192.

Diaz: What no market for bandwidth? A wave is just bandwidth and when I talk waves I also mean OCXX services as well as gigEs. Two gigEs fit on an OC48 and 8 fit on a 192.

Freedman: What I'm saying is, no one's buying lit or lambda services now; certainly not 80 customers worth.

Diaz: Wow really, sorry Avi. I know a bunch. I know people buying it rather then dark fiber since they can buy OC48s at one year contracts.

Freedman: Again, I disagree. I'm sure there are 5 or 10 but not enough for Joe Network (or probably even Cox) to go out and sell 40-80. Then again, equipment's cheap but even so it is a ROUGH market to sell lit or semi-lit long-haul capacity now.

Doncaster: If nobody is buying them now, I guess the price hasn't bottomed yet. If you're a carrier that has cleaned your balance sheet through chapter 11, as long as you cover opex then you're still in the money. Which leads me to a question that will help me understand better the factors in bandwidth pricing: On a short route like NYC-DC where no OEO is necessary, and perhaps even no Raman or EDFA amplification is required along the way, what would the monthly opex costs be for fiber lit up with 80 OC192 lambdas?

Diaz: You will definitely have to re-Amplify on NYC-DC especially if you want full use of all the channels. ReGeneration may not be an issue, while several companies have addressed the tilt issue that occurs.

As prices drop new business uses should occur as new applications develop. Also many companies will decide to build on their own. Can we also say government? I've also been seeing it going international down here.

Freedman: I know a bunch of companies (more than 5) that are finding a hard time getting one customer every few months for cheap inter-city long-haul (trying gov, enterprise/financial, ISPs all). But maybe someone's selling and it's just not my buddies :)

Doncaster; Cheap would be less than $3Kper month for Gig-E from NYC-DC. I haven't seen anyone in that price range, but I have seen less than $15K/mth for OC192 on the same route. The most annoying thing about shopping for long-haul circuits in the OC3 to gig-E range is the sales reps who think they're offering you a great deal for an unprotected OC3 @ .020/DS0-mile. I've been shopping around for an OC3 to Seattle from Toronto or NYC. The way things are going, sometime next year pricing should drop down to my target of <$3K/mth.

Freedman: Also, does anyone actually know of any carrier REALLY selling lambdas, vs. an OC192 circuit? Most I know won't sell a lambda because most gear that takes lambdas doesn't have filters to prevent someone stomping on other frequencies with bad or mis-configured gear.

Diaz: My mistake. Technically Avi is right. I use them interchangeably because I deal with the public and they do. Yes I am speaking 1310 nm handoff to customers, and they do the conversion onto the fiber at whatever wavelength. Better?

Freedman: Yes. I wasn't trying to bust on anyone, but it was most ambiguous as the question was phrased below. And it's a question I'm actually curious about, given the marketing hype about 'selling lambdas'.

Doncaster: [Regarding Avi's question about REALLY selling lambdas], — yes, 360 sells lambdas. You don't get the actual wave though, so from that perspective the answer would be no. The difference between their wave and OCx pricing is that for wave you get plugged into a transponder. For OCx you plug into one of their ADM's. Their list price for an unprotected OC48 is .0015/DS0-mile. A 2.5G wave is 1/3rd less at .0010. With that OC48 they can reroute your path with a software reconfiguration on their equipment. With the wave I believe it is less flexible.

They can also do Gig-E "waves" with their Optera 1600 gear, but they don't offer it to their customers. Hopefully I can get them to recognize the opportunity and sell me one (I already have an unprotected OC3 from them between Toronto and Manhattan). Gig-E would be a lot cheaper for me than OC12.

Odlyzko: What are the units in the .0015/DS0-mile pricing for an unprotected OC48? Is it in dollars, and over a year or a month?

Doncaster: US dollars per month.

And while we're on the topic of dropping prices, here in Canada the best deals on long-distance voice (using an underlying VoIP network) are at 2 cents per minute (and that's Canadian pennies, not US). That's for continental US/Canada, any time of the day. Just in the last year or so the local ILEC (Bell Canada) has seen LD revenues drop below local phone line revenues.

Odlyzko: Thanks for the clarification. Apropos LD voice, in the US I am told that you can get it (at wholesale) for about 1c/min over the traditional circuit-switched network, including access charges at one end. (That is, if you have your own connection to a long distance company, and don't have to pay access charges to your ILEC, you can call anyplace in the US, and probably Canada, for 1c/min.) Wholesale rates for LD in the US (net of access charges) have been well under 1c/min for at least four years. Thus VoIP is a threat not as a technology per se, obsoleting the traditional network, but rather as a way to bypass access charges and all the marketing/billing/... costs of the LD carriers.

Doncaster: I'm talking retail here. C$10/mth gets you 500 minutes of LD. C$20/mth gets you 1000 minutes of LD.

[On a different subject:] My belief is that having a POP in a place like 60 Hudson (or 25 Broadway or ...) is cheaper insurance against transit price hikes - if your current transit provider hikes up prices, switching to another is just a cross-connect away.

David Diaz: [I disagree with the fear of increases]. Laser speeds are increasing and the number of channels is also. Then we are using them more efficiently with new equipment that is actually cheaper. Not to mention cheaper maintenance and fewer people. Now what I mean by new features is that you used to have to drop a whole wave at the Z location, but with switching technology built in, you can now drop parts of a wave in several Z locations. More efficiency. Or the price will rise and then the bankers will finance again. The rules of a dwindling resource apply here, you know the R's. Right now with an abundance, we will all be a bit sloppy.

Freedman: [Interconnects are another cost factor.] If people do peer widely they have a lot of relatively high $/mb pipes until they have enough traffic on them. In the 25-50/mb range minimum is the current range for interconnect average costs, though that is coming down as people centralize in peering centers and add private peering - again, if they are at least at the few gigs per second level.

If people are just in the US, and absorb a relatively high fixed cost to go to all of the super-big-boy exchanges, maybe that doesn't apply, but then again it isn't clear that such exchanges will become the core.

COOK Report: What Avi says above is that people who peer widely still have to pay exchange points to interconnect with their peers there. Also if they have smaller levels of bandwidth spread out among a large number of peers, those levels are more expensive Right?

Woodcock: Right. And, more importantly at most exchanges, the backhaul from the exchange into the core of your service-provision network. As you asked later in your message. So these are all relatively high step-function costs... That is, the granularity in which you have to purchase those services is very coarse. You have to pay quite a bit for the minimum degree of participation, even if you have very little peering. That's the problem that a lot of big Asian ISPs get into at the PAIX, initially. Huge expensive pipes across the Pacific, which sit pretty empty for a month or two or more while they negotiate peering.

James **Spenceley** (Australia) Certainly the initial pipe this is case, however with the opening up of the IRU market, purchasing options have allowed providers to have dual un-protected capacity, facilitating running these at greater than 50% utilization each, allowing for better utilization for the next pair as they are turned on and only reducing redundancy slightly. Of course that doesn't help people with getting peering, but very few wouldn't purchase their transit our of the same locations as they peer.

In our case we needed to have a US POP, so it was more cost effective to place it in a IX than in XYZ co-lo. So if planned well, avoiding such anomalies as expensive local loops and aggressive IX operators, it can be just as cost effective to get a little peering than none at all.

Freedman: Yes [as to Gordon's and Avi's points about peering economics above.] For example, at ATM exchange points if the OC12 costs $15k/mo (assume that's the only cost), and you can use 300mb on a 95th percentile equivalent (of the 500mb-ish available) with 450-500mb peaks, then the minimum cost of traffic over that ATM point will be $50/mb, not including equipment, people time, rack costs, or layer 1/2 transit to the XP if needed.

For those not doing huge amounts of traffic, going to an Equinix type facility costs $10k/mo-20k/mo between racks, loops in, and Equinix costs. Plus 250-500/mo per peer for private interconnects, I think. So if you do 1gb+ at that XP the cost can come down per mb but most of the networks not in the top 10 probably are still looking a at 25/mb or so cost before backbone, gear, and people costs, I'd think.

COOK Report: A question on transit prices - $100 per megabit at rates of oc12 (644mbits) was mentioned as a floor. At OC48 is it still $100 per megabit?

Freedman: Generally yes. Though there are a few larger providers of questionable state who will do 50-60/mb at OC12 commit levels.

Woodcock: Often directionality of the traffic will make a bigger price difference than volume of purchase. SO there are plenty of $70 deals out there right now, if you can guarantee unidirectional traffic. That is, the deal is $70 in one direction, and full retail in the other direction.

COOK Report: Supposing one of the "big seven" decided to drop prices on OC48s to $50 per megabit? Could the other seven absorb the increased traffic they would receive without significant cost because the backbones are lightly utilized? Or would no member of the oligopoly dare do such a thing because then the pressure would be there for everyone to drop prices to avoid loosing market share.

Freedman: They could and would do it if their product management people decide to, in my opinion. Whether the provider's interconnects would congest because of rapid movement of traffic is a good question, but likely the people in the "tree house" would increase capacity.

Woodcock: Actually, I don't think we're plateaued... I think all of them are actually doing this all the time, forcing each other's prices down continuously.

COOK Report: Still what is the average turnover time on transit contracts? A year? 2 years? Six months?

Freedman; Four to five years was the norm in 1995 in many cases. 1-2 years is the norm now I'd think. Most people sign 1 year. Those who think BW prices will go up do 2. Few do greater than 3.

Woodcock: Nearly all retail contracts are 12, 24, or 36 months, with the vast majority being 12-month. A very few very large customers have gotten that down to 1 month or 2-6 months, by buying in large volume from multiple providers _at an exchange_ where the providers know that there's no barrier to the customer swiitching.

Odlyzko: Now the assumption that pervades the discussion on this list is that costs of peering and transit dominate. In particular, Gordon in his Introduction writes about "the technology director of a large CLEC [who] said: "that sounds pretty much like how we model and our long distance circuit interconnections with the PSTN. Do it right and you are profitable. Do it wrong and you are history." I can well believe that this used to be the case for voice connections, but would like to see some real data for what goes on today. It was surely also correct a few years ago for ISPs, when transit was very expensive. But is it true today? I would like to see some real data. My strong impression is that while the costs of transit have plummeted, the prices paid by the end users (at the T1 and sub-T1 level, where the bulk of the revenues are) have declined much less, and of course there is a far greater density of them.

Woodcock: That's certainly my impression as well. Bulk transit costs seem to have come down from perhaps $800 megabit to $100 megabit over the last four years, while retail has come down from about $1200 to about $550 over the same period. Does that square with other folk's general impression of the numbers?

I wouldn't argue that that indicates greater profitability, though, by any means... Just that people are dumping excess capacity to big customers who can chew it up quickly, at whatever they can get for it.

Costs Move from the Center to the Edges

Odlyzko: I was certainly not implying greater profitability, although that may also be true. This is probably the result of the natural evolution of the industry, with costs moving towards the edges. We can see something very similar in the computer industry. The power of the leading-edge microprocessors has been increasing for several decades at about 60% a year, as described by Moore's Law. Their prices have stayed stable, at a few hundred dollars each. On the other hand, the prices and computing power of 5 MHz 8-bit embedded microprocessors have not changed all that much.

Freedman: I think retail has come to $180/mb to $350/mb or so, but not $550/mb. UUNET will quote you $400 per month for a t1 (full) or $225/mb for collocation, I think.

Hussain: I mostly hear much lower numbers for paid peering between $50-100 per meg. It varies obviously depending on who is selling but numbers as low as $25 per meg for paid peering are rumored though between $50 and $100 seems more the norm. These deals may be cut with reciprocal arrangements elsewhere for purchase of capacity. It would be really valuable to have some trans-

parency in these numbers but I fear that is unlikely to happen.

Nowlin: Paid peering often takes the form of one party purchasing all the loops/cross-connects to the other party. In extreme cases (1998 time frame) there were cards for routers and even fiber discounts in exchange for peering. Still paid, but more difficult to pin a dollar per megabit charge on. Loop or cross-connect MRC covered by one party of the two peers is still the easiest way to deal with the downward spiral of per megabit charges when offering up a middle ground to peer/no peer negotiations.

Odlyzko: Along the same lines, transit revenues from ISPs that can get the $150/Mbps per month prices from Tier 1s simply do not come to all that much. Most of the money the Tier 1s make comes from end users. That should be kept in mind when evaluating the tactics those guys use in setting their peering policies.

There are some other basic assumptions in this discussion that I have doubts about. For example, in his first interview with Gordon, Bill said "It probably doesn't change things. My point is that if you are buying transit from two providers and you are adding your costs and your profit margin and reselling the result, you price to customers is going to be higher than their price to their customers. And your customer might as well bypass you and go directly to them. You have no value add."

Is that really so?

Weller: Correct. Correct again. While OC3+ pricing from Tier 1s are at $150/M/month, the cost of single T1 bandwidth is more like $250/M. That alone is a measurable margin. But also consider that aggregation exists at the wholesale level, such that for every T1 sold to a customer, you really need less than 300K to cover it.

Woodcock: My point wasn't so much that resellers can't carve out a niche for themselves anymore, but that wholesalers have _always_ thought of themselves as being able to compete directly with their down-stream resellers for the same customers. That's _always_ left resellers with a difficult job of customer retention.

Odlyzko: Yes, indeed, and that shows again where the costs are. If you sell T1s at $250/M, and can statistically overbook the transit bandwidth, you are paying $150/M, and collecting $1,250/M. (Actually, probably even more, since the $250/M price for a T1 seems exceptionally low. I typically hear prices closer to $1,000 per month for a T1, which works out to $660/M, but perhaps my numbers are off.) In any case, most of the costs are local.

Woodcock: Cheap T1s have been $600/month for a long, long time. Expensive T1s have come down from the mid-$3K range to the $900 range over the time I was selling them.

Odlyzko: If you are buying transit from two providers in the OC3 capacity, say, and your customers are buying T1s from you, you are providing real service (aggregation of traffic, as well as all the hand-holding for the customers, etc., which is where most of the costs are).

Woodcock: Mmm, you're right, and I was guilty of rhetorical overstatement... My point was basically that if one is simply reselling transit, rather than peering one's self, one is pretty vulnerable to one's transit providers setting one's prices for one. Not, as demonstrated by the relative change in wholesale and retail prices over time, that that appears to actually be happening. I guess I resist looking at that as a basis for long-term business decisions, since it seems like a bubble.

Odlyzko: An oil refinery (not part of an integrated business) sells to wholesalers, who sell to gas stations. This is very common, and all stages have value add. This is especially true on the Internet, where there has been little successful vertical integration. Bill talked about increasing difficulty in ISPs differentiating themselves on the basis of quality. However, we still see large differences in prices (about a factor of 2 in prices for T1s, I hear), which shows that Internet connectivity is not a commodity.

Woodcock: Ah, but that's down from a five-fold difference maybe six years ago.

Odlyzko: Interestingly enough, if you look at the STM-1/OC3 lease prices on major US and European routes in 1Q2002, shown in

<http://www.telegeography.com/resources/statistics/bandwidth/tb02_lease_prices.html>

you still find more than a 3-fold spread from low to high.

Tier 1 Versus Transit Free

Ralph **Doncaster** reacting to the interview with Bill **Woodcock**: It might help if you define your version of "tier-1". The generally accepted definition seems to be "transit free", which would include AboveNet/MFN, Verio, Teleglobe, and Gblx as far as I'm aware, and maybe a couple more.

Woodcook: Yep. Sprint, C&WUSA, WCom... I just had an interesting conversation with Tom Vest, at AOL, who believes there are six transit-free ISPs. AOL has, for obvious reasons, been doing a pretty careful analysis.

Doncaster: I meant in addition to your list of Sprint, C&W, ... sorry for the confusion. However if there are only 6 transit-free networks, that would be a VERY interesting piece of information. I'm guessing that would mean there's a lot more paid peering relationships among some of the big networks than people would think. Since full transit relationships are easy to determine from places like route-views, but paid peering looks just like settlement-free peering, then any information about who are the 6 that are truly transit-free would be quite valuable.

COOK Report: My understanding of Farooq's definition of Tier 1 is that it equals Bill's donut hole.... those networks that

peer only with each other and with no one else. But from what Bill says below his donut hole and Farooq's internet core (Big 7) don't seem to be identical. Anyone willing to comment on the difference between the Internet Core of 7 and transit free??

Woodcock: I think we're all basically getting to the same two things. There's the transit-free thing, and there's the don't-want-to-peer-with-folks thing, the two of which combine to form what I've been describing as a donut-hole, and which would generally be thought of as the "tier 1" or "default-free" part of the Internet. I wouldn't say that there's no peering across that border, just that there's markedly less, or sufficiently less as to constitute a useful distinction.

Hussain: I agree with Bill's statement above. I also think it isn't necessarily helpful to focus the don't-want-to-peer mndset too exclusively on the networks that were involved in the I-Core initiative. It's their apparent behavior and not the principle that they find it efficient to directly interconnect at IX's that I find to be the problem. IF they were at multiple IX's and that they would be willing to sell transit and be open to peer with others there THAT in my view would make a whole lot of sense.

I feel that they believe to do so would push the floor of transit pricing into the underground car park let alone the basement. It's not clear to me that this would necessarily happen but I'm confident that this is their main concern. Their business models depend on transit revenue and they seem to have gone into a wagon circle to defend it. Such approaches in my view generally don't pan out successfully.

Smith: The US marketplace is more of a mystery to those outside because of the "Tier 1" mystique.

Freedman: Agreed. DT, FT, BT, Telstra, KDD, etc are all "Tier 1"s marketing-wise in their home areas (not just countries) but not transit-free in the Internet sense.

Smith: The reason for differences between the marketing Tier 1s and those ISPs who have to pay no one for transit is clear - marketing is all powerful. Here in the AP region we have many of the US ISPs selling services and claiming Tier 1 status, but completely unwilling to back-up this claimed status with any sort of proof. But then many people can be completely bowled over by impressive customer lists, quotes of big backbone bandwidths, impressive lists of peers and interconnects, and the magic words "we connect to UUNET in N places".

As far as I'm concerned, the ISPs who pay no one for transit make up the Internet core. The "Tier 1s" are those who have really good marketing departments. The latter list will be quite long - I believe the former list would be hard to verify without examining the deep business relationships inside every major ISP. And of course there will be an intersection.

Furthermore, it is still possible to be a paying transit ISP, and be one of the largest around. How do you measure size though? Through backbone capacity? (And if that, how do you account for over engineered backbones, under engineered backbones, over provisioned capacity, etc...) Through prefixes announced? Through ASes transited? Through address span announced? (You can get some idea of what I mean from www.cidr-report.net - are those top ISPs there really Tier-1s, or are they in the free-transit zone? Or are they something else?)

Freedman: Just went to www.cidr-report.org. While someone could look at BGPand do a 'routing tier 1'type analysis (which still wouldn't show paid vs. free transit), number of routes announced has never been a good indicator of Tier 1/ core network/transit free or not.

COOK Report: Paid peering again is the purchase from network x of the delivery of all packets that terminate in network x, but with network x having no obligation to deliver your packets elsewhere? Do I understand that correctly?

Smith: That is paid private peering. And as has been mentioned before, the other option is paid transit peering where network x has an obligation to deliver packets somewhere else. That "somewhere else" depends on what you pay network x.

COOK Report: When a network says it is transit free, is there any way to verify that?

Freedman: There are some ways to verify it, primarily from looking at BGP. I suppose typing in the AS in the CIDR report and examining the 'upstreams' might help, though I tried a few ASs and found bogus ISPs identified as upstreams.

COOK Report: Speaking of non US centric, are any of you aware of the kinds of negative financial impacts on foreign carriers of the transit situation in the US of which Farooq speaks? If so would you elaborate?

Woodcock: Well, the ability to pick up very cheap transit at some of the large European and US exchanges isn't helping growth at AP-region exchanges. For instance, it's more practical for some of the Nepali ISPs to get space segments to London than to Singapore, for their next stage of growth, since the difference in the transit which they can also pick up at the exchange is so great.

Klein: See http://peering.adelphiacom.net/ for [our] peering info.

Doncaster: So with a network that big with over 4Gbps of traffic, why not connect to 25 Broadway for NYIIX peering? (Considering that you have a node in Phily, and that an OC48 lambda from Phily to Manhattan is dirt cheap)

The other thing I don't understand is why anyone connects to AADS now. Is it just that people locked into long-term contracts back when paying $7K/mth for a peering OC3 was cheaper than transit?

Klein: Economic changes at Adelphia have made issues in NYC more complex than could be wished. We have a node in Manhattan that has been doing nothing more than listening to routing updates from Philadelphia for six months. Chap-

Economics of Level 3's European Connectivity

Level 3 Makes Excuses

COOK Report: From a source in Europe – Any Comments?

I just got this from Level 3 support after complaining about poor network performance.

Statement 1.

Level 3: "Level 3 is interconnected in multiple locations with UUNET in Europe. This capacity has been fully exhausted in the last 2 weeks as a result of the KPNQ and Teleglobe failures and influx of customers seeking emergency capacity on the Level 3 European network.

Hussain: DANTE, Belgacom, and a bunch of other very large KPN customers went with Telia and others than Level 3. SURFNET the Netherlands R&E network elected to go with Level 3. All the networks that absorbed these larger customers experienced increased traffic volumes and resulting strain on their peering capacity - especially as there were two less peers to exchange with.

It's been widely reported that the shut down of E Bone happened pretty much without a serious disruption for many though not all. So the statement that the KPN and Teleglobe shutdown resulted in Level 3 picking up all their customers causing congestion on their links to UUNET I doubt would bear up to any scrutiny. Note that Level 3 is reporting these current problems as attributable to KPN and Teleglobe some three months after those networks were shutdown. This makes it seem even more unlikely that this a reasonable explanation of the problem.

Level 3: If we had left the UUNET European links in place, we would have been running them beyond 100% capacity at this time and would certainly be dropping customer packets. Since we strive to never drop packets, we looked for the next best routing to UUNET with sufficient capacity to handle the demand.

Hussain: Perhaps Level 3 cannot afford to continue to maintain the capacities required with UUNET in Europe. So given transit pricing it may be cost saving measure on Level 3's part. I doubt that UUNET would be interested in upping the capacity of their peering links with Level 3 also.

Nowlin: I heard around the grape vine that credit issues held up the installation of proper capacity on the UUNETside in Europe. Loops were purchased in rounds and it was UUNET's turn. Level 3 had capacity to handle the customer transitions from KPNQwest/Ebone/GTS, even with the AOLtraffic they carry in Europe. This was from a third party but believable given the pattern of loop rounds to augment existing peer capacity has been in practice for several years. Adlex tools let me know that AS702 was being seen behind AS1239 during that time as well. That would indicate some AS701 congestion was being worked around. AS7176 was behind AS1239 during that time as well vs. AS1.

Hussain: Basically, I've had the same confirmation. UUNET were not able to implement a planned upgrade of capacity because they're not allowed to spend any money under Ch. 11 without explicit approval of their creditors. Level 3 have had to route a lot of traffic through the US in these circumstances but nobody feels it's a result of the KPN [Surfnet] and Teleglobe customers that were won by Level 3 as they claimed in their release. As those networks shut down the customers were spread out amongst multiple providers.

Level 3: Since Level 3 is one of the largest Tier 1 providers in North America today, the largest interconnection we have with UUNET is our peering on the East coast of the US. So, we have temporarily suspended sending traffic over the smaller European links in favor of these larger North American interconnects, until we can resolve the European capacity.

Hussain: Cost saving again?

Level 3: The situation is being worked on currently, but we estimate it may take some weeks to return UUNET based traffic to the European Interconnects and to have them upgraded to handle the sudden increase in demand. Until this time, customers can expect to see increased latency to AS 702, but we believe this is preferable to experiencing packet loss in the short term.

Level 3: We apologize for any inconvenience of service degradation experienced during this time of crisis across the European Internet, as a result of KPNQ and Teleglobe going into liquidation."

Hussain: Meaningless waffle.

Statement 2.

Level 3: "The demise of the KPNQ and Teleglobe networks continues to strain European interconnectivity. This is not a Level 3 specific problem, but rather an industry wide issue. While the current traffic exchange with UUNet is not optimal, Level 3 is fortunate enough to have significant bandwidth to UUNet in North America to absorb the load. Level 3 is working diligently with UUNet to reach a resolution as quickly as possible. While we are unable to provide you with specific dates for resolution, we expect to be able to provide you with more detail in the coming week."

Hussain: I just have a lot of problems believing that Level 3's problems are to do with KPN and Teleglobe's liquidation.

ter 11 turns bean counters, lawyers, and judges into network engineers. :-$ At some time over the next 90-120 days I hope to announce additional peer locations. Can't say much more.

Traffic Patterns – How Much is Local?

Jere **Retzer** on September 10: Some questions for the list:

1. I've been of the opinion that as real time interactive applications like VOIP and video conferences gain greater market penetration that QOS between networks will become important. If true, how do we get there? Do we have a chicken and egg thing that the applications will not take off until the quality is there but the quality will not develop because there is no demonstrated market?

2. I've also maintained that these applications will drive more traffic to the local area vs the global Internet because people tend to interact more with people with whom they have an ongoing relationship. Telecommuting and telemedicine are examples of applications that I would think will connect mainly people who are less than 200 miles apart. My hopes are that these applications will finally begin to take off now that we are seeing decent broadband last mile penetration. If right, this would seem to argue for high quality, reliable links between local/regional providers Å\ potentially via a metro/regional exchange. Does, or should this change the view that peering should always be as cheap as possible and that reliability does not matter? The example I always give is that from my desk at Oregon Health & Science University to a local community college just five miles away used to be as many as 29 router hops via Sacramento, San Jose, San Francisco, and Seattle (Everett) and! that there was no way to ensure decent quality. We've changed that with a direct interconnection.

Questions:

a. Are we seeing any evidence of a traffic shift back to the local/regional level or does it seem likely that this will occur?

b. Is there an important role for high quality connections between local/regional networks?

3. Given the need to reduce local access costs, do metro and regional exchanges make sense as a means to consolidate circuits and in this way to effectively bypass the incumbent carriers? If so, does this demand higher reliability and quality? It seems to me that using exchanges in this manner could be a very important survival strategy for competitive providers Å\ simply cannot afford to trench out to every small-medium customer. At one point, during the height of irrational exuberance we had 36 franchises digging up Portland streets to get to the same large customers. That clearly was not going to work from a business perspective.

Given the natural evolution of high-end applications from large to small customers it would seem that these small-medium customers will soon be demanding more advanced capabilities so there should be a market for relatively high-speed circuits. How/who will build these given the capital crunch? I don't see small-medium companies buying high-speed circuits to all their business partners the way a large business might to their branch offices and I don't yet see the Internet providing the reliability or quality needed for the applications. Therefore, I think the answer may be high quality 'local Internet' or Intranet access. Does this seem like a viable concept?

4. What is the best way to convince ISPs to offer transit via a metro/regional exchange? Most are used to thinking of exchanges as simply a place to peer. How does a metro exchange make the leap from peering to transit exchange? Are there any special considerations to successfully package an exchange for both peering and transit? We may soon expand to couple more data centers linked to our original location via GE over dark fiber. Seems like it would work well if we could get some providers to offer transit at each location and then enable peering between. Thoughts on this approach?

Odlyzko: Very good questions. Locality of traffic is something I have been interested in for quite a while. It occurs in all communication services that I know of, other than possibly the Internet, and there are even so-called "gravity" models that describe it quantitatively. It is natural to expect the Internet to move towards that model as well. Does anyone have data that could be used to support this thesis? (The growth of peering in Europe and Asia, as replacement for peering in the US, is part of this trend.)

Lindqvist: While at KPNQwest I was looking at this. What we saw was that countries that had a "relationship" (language or culturally) had a clear shift in more "regional" traffic from a very US centric traffic. What I saw (this is around a year ago) was that there where several countries that where above 50% of their traffic staying national, and around 30% of the traffic staying in the region, with the remaining traffic going elsewhere. Just two-three years ago, well over 50% would have been US traffic.

I discussed this with other major European networks and they seemed to have seen the same trends.

Retzer: Some [additional] comments:

I really like the description of a single peering fabric in a metro area with multiple transit exchanges. That seems to already be evolving in Seattle.

Thanks for the consolidated exchange list on PCH.

Woodcock: We don't have any _known_ duplicates. It's often difficult to establish what's a duplicate listing until sufficient distinguishing information is gained for different entries. We retain all defunct and non-exchange entries to avoid them being re-listed again in the future. And we collect historical information about them, if it comes our way.

Retzer: This is a good reference although I think there is a lot of duplicates

and a fair number of dead entries. We're listed twice.

Woodcock: My understanding was that you had a v4 uni-exchange on 198.32.195.0/24, and a v4 multi-exchange on 198.32.196.0/24. Is that not the case? If not, the 196.0/24 block should be returned for reallocation to someone who can actually use it.

Retzer: PAIX-Seattle is listed three times, for example.

Woodcock: v4 uni A, B, and multi.

Retzer: How do we go about getting a couple of the PCH looking glass servers?

Woodcock: We've got a requirements list that I'll append. Basically we deploy as quickly as we can, prioritizing for geographical distribution and number of peers.

PCH Host-facility Policy and Requirements

PCH maintains a queue of exchange facilities into which we are installing route-server/looking-glass equipment. We have a number of requirements of the host facility, namely:

- 4U of 19 rack space for the equipment, including 15 amps of 110V power

- Four standard computer power cords of appropriate local connector type

- Two 10Base-T or 100Base-T switch ports on the exchange switch fabric

- Two IP addresses on the exchange subnet

- One 10Base-T or 100Base-T interface on a transit network

- A /29 (eight addresses) of globally-reachable address space for remote administrative access to our equipment

If your facility can supply all of these requirements, or some reasonable equivalent, we can begin the process of prioritizing your facility for an installation. Priority is generally given to facilities which host large numbers of Internet providers and to facilities in developing nations. Unfortunately our resources are limited so facilities which can cover our costs for installation (currently US$7500) will also receive higher priority.

Retzer: OK, thanks. That explains your methodology. It works and is a valuable resource. The reason I asked is because I'm wondering if local peering exchanges are beginning to disappear in the US due to the collapse of transit prices. I did a market evaluation about 18 months ago and it seemed like there were more active local peering exchanges then in the US than now. This folds back to the business model discussion.

Woodcock: If you know of any that have actually gone defunct, I of course need to know about it, to update the spreadsheet. The only one I can think of that actually went away is MAE-East FDDI. MAE-West has certainly been shrinking, but others seem to be fine.

Retzer: We have a redundant setup with 195.0/24 on one switch and 196/0/24 on the other.

Concerning transit across exchanges, the question has come up here about tools and methods for accounting for transit to multiple customers and/or combination transit and peering to the same customer via a single port. In the latter case, the contract would peer some local addresses but provide transit for others. I've heard it argued that most ISPs expect to have one transit customer per port, period. End of story. Are they just being lazy or is this really a problem for some platforms? What are most using to account for the bits in this situation?

Kurt Erik **Lindqvist**: At KQ we through history (from around 1996 until May this year when KQ went belly up) tried to achieve advanced accounting based on

- volume (total) - destination and source - Sampling of any of the above.

The last attempt comes from the fact that at today's interface speeds in the backbone, the data generated is simply to much to do anything useful with, as well as vendors not supporting anything else.

As an exchange point operator, I would say that if a ISPwants to sell any form of usage based service - it's their problem. There are solutions if the interface is not high-volume.

Woodcock: When I see two ISPs peering and selling each other back-up transit, they typically either use vlan tagging or frame pvcs to create three pseudowires (and thus three _logical_ interfaces), and peer across one, while each offers the other transit across one of the others. That way you can keep track of who was intending to use which kind of traffic.

Retzer: So set up one VLAN for each transit agreement and one for each peering agreement and use mac-based accounting?

VLAN 801 Å\ general peering VLAN 711 Å\ transit to A VLAN 712 Å\ peering to A VLAN 713Å\ transit to B

Doncaster: When you receive a packet destined to A, you're going to send it over VLAN 711. Besides that, why would a transit provider keep a settlement-free peering arrangement with a customer?

Freedman: Because some people get idealistic about this stuff :) "We should exchange traffic for free but I'll pay you for access to YOUR peers/transit". So why not make some money on it?

Transit Accounting and Billing Issues

[Earlier] **Doncaster**: Cisco mac-accounting would do the trick here. I use it for some of my customers where they are all on the same Ethernet segment.

Lindqvist: Have you tried this on a fully loaded GigE interface? Doesn't scale very well and the granularity is rough. How do you know what of the traffic is peering and what is transit?

Doncaster: I can count on my fingers the number of providers in Canada that can fill a GigE with transit/peering traffic - I'm not one of them — yet ;-) This wouldn't work (as you note) if you want to peer and do transit for the same AS.

Stuart (on September 12): A certain router vendor did not support SNMP counters on 802.1q sub-interfaces until very recently.

The issue for one-customer-per-interface, though, is often that isolation of customers from each other is good. If you don't get that up front, then having to pay out on an SLA violation to customer A because customers A and B share fate at layer 2 and customer B did something stupid should bring the practical aspects of that advice home pretty quick.

Fronting an 802.1q router port with an aggregation switch implementing per-customer VLANs, and using the switch ports for accounting, was popular for a while for (a) keeping customers from seeing each other at layer 2, and (b) avoiding the expense of high-cost router ports just to get per-customer SNMP counters on a device that could be easily correlated to something measured on the customer's device.

The first bit prevents customers from being billed due to other customers'multicast or broadcast traffic showing up on their dollar-meter. Billing disputes cost you (money in the form of staff time and credits, and good will because you can't measure usage fairly), and engineering should strive to minimize them.

That last bit is the second reason that having some device, logical or physical, on which per-customer SNMP counters calculate usage are good. If the customer is also calculating usage, it's probably by measuring if InOctets/ifOutOctets or if HCInOctets/ifHCOutOctets on some interface and comparing the numbers measured to yours - with a certain threshold for "sameness" in mind, beyond which you get a billing dispute. You can substitute some other means of measurement, the point is that a means of independently verifying usage by measuring some other device besides yours can be very helpful in maintaining customer happiness.

Freedman: One of the interesting challenges is how to sell transit but be willing to keep peering. In the scenario you described, the party buying transit gets free outbound but not inbound.

For example Netaxs at one point bought transit from Abovenet. At the time, Netaxs kept its Abovenet transit. So Netaxs got to send traffic to Abovenet's customers for free - but ALL traffic from Abovenet's customers and from Abovenet's peers (that's the transit part) went to Netaxs over the paid pipe. Now, Netaxs sent more to Abovenet than it received on the transit pipe so it was no big deal.

But among enlightened parties, this is still an issue with no good technical solution that I'm aware of. I suppose you could write perl scripts to configure a juniper to do policy routing based on source address, but just forget it with Ciscos.

And while an enlightened provider might be willing to keep peering up while selling transit, I've not met any who'd give a 'credit' based on Netflow stats.

Woodcock: Ideal case:

A sells transit to B on a pseudowire we'll call AtransB.

Doncaster: So how do you implement this without chewing up all your router CPU time? On my Cisco 7206 turning on a route-map to implement source-based routing nearly triples the CPU load.

Stuart: One possibility: on Cisco, UTI tunnels; on Juniper, CCC/TCC.

Woodcock: Across that connection, A sends full routes to B. B sends customer routes to A. B may send any traffic it likes, but in practice only stuff to folks other than A's other customers. A must re-advertise B's customer routes globally. Any traffic which A receives destined for B across its own transit links it can put through this connection. Any traffic which A receives destined for B across its peering links it can put through this connection.

A and B peer across a pseudowire we'll call ApeerB. Across that connection, each party advertises only customer routes to the other. Each re-advertises routes learned across this connection only to their own customers. Any traffic A receives destined for B from its other customers it must put across this link, and vice-versa.

B sells transit to A on a pseudowire called BtransA, which is the mirror-image of AtransB.

Each party just needs to keep a packet filter on their own side of the peering link to make sure the other side doesn't attempt to dump non-peering traffic in that way. And each party might want to keep track of who the others' customers are to occasionally compare netflow to interface counters to make sure that the other side isn't dumping peering traffic in through the transit link.

Stuart: I would argue that the ideal case is configurable counters in both directions, of the style provided by counting inbound traffic destination addresses using destination-based accounting, where inbound traffic is differentiated by destination address and outbound traffic is differentiated by source address. That would let you slice and dice traffic up to the maximum depth of unique counters provided by the implementation, where you would presumably create different buckets based on cost of goods sold (COGS) for each bucket.

The simplest case would be two buckets in each direction: did the traffic {come from,go to} another customer (for which there is one costs basis), or did it {come from, go to} a peer (for which there is a different, typically higher, cost basis). Add a third bucket, and you can split peers into for-fee versus settlement-free.

All the schemes I've seen to serve a customer with multiple connections where different routes are announced/received on each connection impress me as being solely for the purpose of accommodating the lack of really robust accounting features in the router. As someone once said, though, you can forward or you can count, but not both. That was pre-ASIC, though; maybe there's sufficient foundation to do something like this now.

Freedman: What we did at Abovenet when I was there was to say "Look, you agree that AboveNet is less than 5% of the Internet, right? - OK, we'll give you a 5% discount on transit if we send more to you than we pull, and if we become a bigger part of the Internet we'll discuss it again." Since this was mostly for Asia-Pac incumbent telcos, it worked. This was Dave Rand's strategy; I claim no credit, though (though we both strove to sell transit and still maintain peering with people).

Stuart: More recently (for those not familiar with the chronology, my turn at AboveNet came after Avi's), we were trying to get something together with destination-based accounting to provide differentiated pricing for traffic sent to us by customers. In that case, routing updates map destinations to buckets based on the presence of communities distinguishing "internal" versus "external," and interfaces have counters attached to them that increment when traffic received matches a bucket.

Freedman: I've heard that AS accounting is what L3 uses, and that it's fairly accurate vs. Netflow.

Stuart: For billing, destination-based accounting provides a mechanism for differentiating received traffic so that different buckets could be billed at different rates.

Retzer: Maybe I'm missing something, but these seem like important problems to solve for 'transit' exchanges or combined 'transit/peering' exchanges to succeed. For IP backbone network "A" (arbitrary letter) to connect to a layer 2 exchange switch in order to sell transit to several local ISPs and other customers, they are going to need some method to account for how many bits they are selling to each customer.

Stuart: As was discussed, VLAN tagging - in the switch it provides isolation to ensure that traffic from other parties on the switch does not intrude, and on the attached routers' physical interfaces it provides a means to associate the traffic with a logical interface where SNMP counters can be polled to determine usage.

Retzer: Closely related, particularly to your example, it would also be ideal to somehow get backbone "A" to peer their "local" addresses to other exchange members, selling transit to the rest of their address space. The routing seems do-able but are there reasonable methods to provide the accounting in order to pay the bills?

Stuart: That depends entirely on the business model of the backbone in question; being able to cleanly announce what is local versus what is not is not always easy.

Destination-based accounting, also already discussed, provides one mechanism for doing the measurement required for billing. Netflow is another possible means to do the measurement.

[On another subject]: An interesting application noted by Vijay Gill was as a detector of transit theft - if you turn on destination-based accounting, configured appropriately, on an interface over which you expect to receive traffic only to you or your customers ("internal") and your external buckets increment, then you may have some issues to address.

Freedman: Of course, AsiaPac incumbent telcos are/were SSDCs (Same Sh.. Different Continents).

THEM: "We are upset. ATT won't peer with us. They want us to build a US network, and even then they may not peer with us." ME: "Hmm. How many people do you peer with in {Korea, Japan, HK, Australia, ...}?" THEM: "That's different. They're not peers."

Hmm. Things may change continents but they don't change much otherwise :)

Doncaster: I've noticed that tier-2/3 cities seem to be getting fewer choices in transit providers vs the big cities - even more than population differences would account for.

I believe a significant reason for this is the change to a wholesale model where an ISP buys dial and/or DSL ports throughout a region, with a single interconnect to the wholesaler. So if the wholesaler is back-hauling traffic from Troy, NYto Manhattan, then the ISPonly has to buy transit in Manhattan. I'd say this doesn't bode well for exchanges in smaller cities (under 1-2 million population).

Lindqvist: I think you are right in your assumptions that the number of transit providers are going down. I personally think this has more to do with the current state of the economy (which has replaced El'Nino as the common culprit) than anything else. There is probably also fewer customers and there is fewer IPT providers in general.

As for the exchanges, I don't see that the number of IP Transit operators would have any influence over this at all. What we are seeing in Europe and Sweden in particular is that local exchanges are growing. For a number of reasons but mostly to keep traffic local as transport costs are higher than IPT costs. In Stockholm today it seems as IP Transit is a lot cheaper than joining a IX. That in it self will be very interesting. What is really interesting is to see what this will do to the routing and the way people build their networks. I see it as a potential threat to stability.

I've been trying to find out what it costs for paid peering, so far with no luck. I know most of the tier-1's do it, so there are people that know. So far I've had one carrier say, "Yes, we can do it. Make me an offer for X mbps of peering traffic, and maybe I'll accept it"

What kind of discount versus full transit could I expect from a tier-1 for paid peering? If I could get a 50% discount I figure it's worth pursuing. Less than that and I'm better off buying transit from a couple of the lower-priced providers.

Pricing and Backbones – Where's the Industry Headed?

On September 12, **Odlyzko**: It is hard to tell how the industry will evolve, but here are some points:

(i) There is a huge excess of fiber right now, enough for quite a few years

Freedman: Or longer on the long-haul, esp. with new gear. In the metro there is never enough fiber.

Odlyzko: Is it that there is not enough fiber, or that it costs too much to hook up to it? There are certainly lots of places with no nearby fiber. On the other hand, many major metropolitan areas have pretty dense meshes of fiber from a variety of providers. However, even then, actual connectivity is not all that great. There appear to be lots of obstructions:

1. The 1000 feet from a fiber ring to the enterprise location is not easy to bridge (it costs a lot of money and involves getting permits to dig up streets, etc.).

Doncaster: The utilization rate on fiber laterals rarely makes economic sense vs copper. On long-haul where you can aggregate 100Gbps of traffic it's a no-brainer. It's only a small percentage of buildings with high-bandwidth customers that can justify the cost of fiber laterals. Most places in North America you can get a copper loop into a customer for less than $50 per month and less than $200 install. For fiber it more than ten times that.

Odlyzko: But copper only takes you to a few Mbps. If you want a DS3 or OC3, you have to go to fiber, as far as I can tell. And that is just the problem, enterprises that want to move into the tens of megabits per second, but are not willing to spend a fortune.

Doncaster: The companies that want more than 10Mbps are rare. Go through your local business phone directory, and ask yourself how many of them need more than 10Mbps. Here in Ottawa (pop. about one million) I'd have a hard time naming 100 businesses that would need it. However I have over 100 business customers paying less than $100 per month for 1.5Mbps ADSL service.

Odlyzko: [following 1. above]: 2. Termination equipment is not that cheap, so unless you have a lot of traffic, you may not be able to afford it.

3. The landlord may want to be involved (there is the matter of space in the basement and the ducts, etc.), which slows things down.

I don't know the relative importance of these factors, but it does seem that it still takes a long time and costs a lot to hook up an enterprise even when there is metro fiber in the vicinity. Does anyone have a good insight into the dynamics of this field?

[Going back to my earlier point:] (ii) Capacity of each strand of fiber is increasing as a result of advances in DWDM

(iii) In addition to excess of fiber, we have an excess of empty conduits, so installing more fiber will be relatively inexpensive (and might be done in the future in preference to lighting old fiber)

(iv) Given the excesses in fiber and conduit, and assuming that no one can monopolize these resources, we can treat the fiber pretty much as a free resource. That means that costs of providing service are in the future going to be dominated not by the giant costs of constructing a national fiber network, but of installing additional capacity, which can be done on a route-by-route basis (new DWDM boxes, then lighting up individual lambdas, etc.) There both capex and opex should be decreasing at a healthy rate, given advances in technology.

(v) Still, Bill does have a very good point. The Internet in its early days benefited tremendously from being able to use the infrastructure of the voice phone network. In the last few years, it benefited from the "irrational exuberance" of the financial markets, which got seduced by the tales of "Internet traffic doubling every three months" into building over a dozen redundant fiber networks. However, Internet revenues are still small. Total Internet service revenues in the US are around $15 billion per year (compared to around $300 billion for the entire telecom industry, including wireless, which is up around $80 billion or so), and, if you accept the calculations I sent around a few weeks ago, transit revenues at the high-bandwidth prices we have been talking about come to less than $2 billion per year. The question is, how can the industry evolve towards a self-sustaining business model? My guess is that instead of VoIPrates going up, pricing will move towards flat-rate access fees. Also, in a year or two, when the current excesses are washed out, price declines are likely to moderate, and match the likely advances in technology, which might produce performance/cost gains of something like 1.5x per year. If traffic continues growing at something like 2x per year, that might produce revenue increases of 20-30% per year, which might lead to a gradual evolution to a healthy environment.

An important point to remember is that core networks are not all that expensive to build. If you look at the costs of build outs of many of the greenfield players, a nationwide network costs on the order of $10 billion or so.

One Hundred Dollar Pricing Floors?

Doncaster: The claims of a $100/mb floors would seem to be 2x higher than reality.

Here's a clip of a post made by kirk@wolf.net to isp-bandwidth a couple

months ago. He's apparently an independent sales agent for a number of carriers. Note OC12 pricing from Sprint & WorldCom work out to under $50/mb. Adding a 0-mile loop in a place like 60 Hudson or 111 8th shouldn't increase that by much more. I've personally been quoted $100/mb for 50mb (95th percentile) burstable IP transit delivered over FE, from QWest in 60 Hudson. Additionally, I have evidence (under NDA unfortunately) that indicates Level3 is selling burstable IP transit to customers buying multiple Gig-E ports for $<40/mb.

OC3 Pricing:

Sprint - $14,600 + Loop (Limited Areas)
NTT / Verio - $8990 + Loop (Cisco Router Provided Free)
Global Crossing - $12,000 + Loop
Genuity - $20,000 + Loop
MCI / WorldCom - $10,000 + Loop

OC12 Pricing:

Sprint - $30,000 + Loop (Very Limited Areas - not for pricing ... for= availability)
NTT / Verio - $24,880 + Loop (Cisco Router Provided Free)
Global Crossing - $32,000 + Loop
Genuity - $42,000 + Loop
MCI / WorldCom - $30,000 + Loop

COOK Report: Any comments on this? Are the carriers so desperate that they have independent sales agents?

Freedman: Yes. Carriers are all desperate.

COOK Report: Ralph, In your above example are you guaranteed 50 megabits burstable to 100 if the bandwidth is there? You are paying 5,000 a month for 50 meg Ottawa New York? Is that really a "deal"? Here the floor appears to be 100 bucks. No?

Doncaster: The QWest quote is $100/mb, yes. The point is that it's for ONLY a 50mbit commit. And the commit means I would commit to paying for 50mbps (95th % measured) even if I didn't use the full 50m. At any time I would be free to burst up to 100M, but if my 95th % was above 50m for the month I'd pay an extra $100/m for the difference. I didn't take the QWest quote though, I went with a better deal that gets me as good (or I think better) quality transit for 1/2 that price.

As for the cost of my OC3 to Canada, it's less than $2,000 per month and it's to Toronto 151 Front St. However that's not paid to the IP transit provider, and I don't HAVE to back-haul the IPto Canada, I could easily sell some of it to potential customers in 60 Hudson.

Freedman: Well, one thing to note - $48/mb at full OC12 rates is really like $100/mb 95th - probably more like $120/mb - on a burstable pipe.

Doncaster: 95th percent compared to the peak on big pipes like OC12 is typically a 70-80% ratio. So 48x1.33 would be a better equivalent, assuming you know how to twiddle with your network to keep peak near 100%. In other words, get the Sprint OC12 and burstable Gig-E from another provider. When the Sprint OC12 nears congestion levels, shift some traffic over to the burstable Gig-E. Now if you aren't good at managing traffic I'd agree that you could expect the 95th% on an OC12 to fall around 300-400Mbits/sec.

Freedman: I disagree, but it could be based on different traffic profiles. Many of the oc12-size pipes I've seen all roughly bill 50-60%ile off of peak, especially if you have event-driven customers, though there are fewer of those nowadays.

And average is almost 50% better than that. But Your Mileage May Vary and that's just what has worked for me. The main point is that $50/mb flat-rate can't be compared to $50/mb 95th percentile without a conversion factor.

Doncaster: I was going by AboveNet's stats. For example look at this rather bursty (and not very full) OC12 http://west-boot.mfnx.net/traffic/sjc3/sjc3-pao1-oc12.html 95th of 226.6 vs peak of 337.6 is 67%. Most of the OC12 stats I've seen with greater than 600Mbit peaks have had 95th in the 450Mbit range. This Abovenet connection to Linx shows a pattern close to this with 95th equivalent to 73% of peak. http://west-boot.mfnx.net/traffic/lhr/linx-1.html

I would agree however that there is no fixed ratio you can count on. I've received quotes for bandwdith based on full pipe, 95th % burstable, as well as sustained average. Trying to compare the equivalent costs of each is rather difficult.

[On September 21]: Despite Global Crossing's financial state, they have a good network. In the same league as WCG, QWest, MFN, etc. I just got a quote from a Gblx agent for burstable IP over OC3 at $77 per megabit. I'm surprised they aren't lower than that.

Someone I know at an east-coast ISP was quoted $80/M with a 100M commit (Gig-E hand-off) by WorldCom a month or so ago. That's the lowest I've heard for a big Tier-1.

Who Dominates — Cable or DSL?

Klein: Cable modem traffic (I would argue) reflects the future image of a broadband-dominated Internet.

At Adelphia we get 30% of our traffic for MSOs, 20% from data centers. If these peer directly this looks like half the Internet to me.

When the RBOCs get relief from current regulations that keep the DSL RBOC networks fragmented into LATAs, we will see a large rise in national networks trafficking in data for residential DSL. This could equal another 30% chunk of the Internet. I then need to subtract traffic from over seas companies like BT, DT, FT and the up and coming Asia-Pacific entities; some who are willing to bring traffic to the US and peer around the tier 1s.

So where does the oligopoly stand when they fall below 25% of the Internet traffic.

Who dominates? I suspect that Cable Franchises and RBOC will have little in common other than a desire to keep local loop monopolies, so collusion is probably out of the picture between these two groups. We are happy to peer with each other.

Odlyzko: I am skeptical. Enterprise traffic appears to dominate (this of course includes people who do their own personal browsing and shopping from home). As far as I can tell, there is simply not enough residential traffic to dominate.

Woodcock: I'd concur, based on time-of-day usage curves from exchanges... Pretty much without exception the workday curve far exceeds the evening curve, both in peak and duration.

Spenceley: The workday curve is also the point at which a large portion of the world's (outside US) retail users are active. Therefore, this should be a consideration. There is a significant amount of aggregated traffic hitting the USA to be exchanged, which shouldn't be overlooked in such estimates.

Woodcock: Sorry, could you elaborate? I'm not quite getting what you're saying here. What I was saying was that, in most exchanges, the workday peak is higher and longer than the evening peak. Are you saying that you don't think that's the case in the U.S. because of AP-region workday traffic hitting during our evening? That's undoubtedly true to some degree, but it doesn't change what I was saying...

Spenceley: Sorry, I was saying the reverse. Asia Pacific nighttime traffic (which in the AP area is almost always the peak) needs to be factored/discounted against the US daytime i.e. its not necessarily 100% _daytime US traffic_ that is accounting for the peak and duration of the you are seeing. I wonder if you discounted this traffic if there would be such a difference between day and nighttime peak ?

Woodcock: At what exchange are you seeing nighttime peaks above daytime peaks? None of the Tokyo exchanges, nor even the Seoul exchanges where you'd expect it, nor Singapore nor Hong Kong.

Spenceley: I have found the source of our confusion. The thread is not about local Asia Pacific IX's or even local traffic, my comments are about traffic leaving from/to those countries hitting the USA and therefore contributing to the USA day peak being higher/longer than the night peak.

If you add up the US<->ASIA links (utilization, not SONET) that would be in the high tens of Gigabits, and might account for why the US daytime peak is either higher *or* (more likely) longer than the nighttime peak. It would be interesting to see if there is a correlation between the extension of the US daytime peak and particular AP time zone residential peaks hitting.

The interesting question is where is this traffic exchanged via Tier1's or at major IX's ? Sadly access to many countries may only be accessible via a 'Tier1'. Peering around them might be possible but in so doing loosing access to most of a countries Internet might be a difficult point to overcome.

Freedman: I agree with Andrew's skepticism [about cable and DSLtraffic dominating] - also, universities have slightly different usage patterns and they are at least as much broadband as desktops at work and as home DSL/cable. A hard problem to see the whole Internet, isn't it?

Nowlin: That depends on your user base... cable modem & DSL is actually quite strong nights & weekends.

Odlyzko: Here is the rough calculation:

Current US Internet backbone traffic: around 100 PT/month (this is about what both RHK and I come up with, perhaps somebody on this list has some other numbers) US residential dial accounts: 50 M average traffic per dial account: under 100 MB/month (if somebody has recent data they can provide, I would be delighted to see it) Therefore total traffic involving dial accounts: under 5 PT/month

US broadband accounts: 15 Million (FCC data had 13 M at year-end 2001, and that is supposed to reach 20 M by year-end 2002, but that is under a very inclusive definition, under which any connection that has at least one direction faster than 200 Kbps qualifies, so I am using 15 M as a rough estimate) average traffic per broadband account: around 1,000 MB/month (I would really like to see some recent data for a large population, as this is an extrapolation based on some older time series) Therefore total traffic involving broadband accounts: around 15 PT/month

Doncaster: The traffic numbers for transit mentioned by Videotron and Rogers indicate higher than that. This was at the peering BOF at Nanog this spring. Videotron and Rogers are 2 cable companies in Canada, with the vast majority of their customers being residential cable users. Together they have almost 1M cable modem users, and avg bandwidth use per user works out to about 3GB/mth.

Rogers recently started re-provisioning all their subscribers from 3M/400K to 1.5M/192K, which may bring their averages down a bit now. Among my own residential ADSL subscribers a substantial amount of traffic is used during the day - about half the level I see during the evening peak.

Odlyzko: Now there is a lot of double counting (peer-to-peer MP3 file swapping between pairs of residential users, etc.), and there is also traffic from other countries, but the general conclusion is that enterprise traffic dominates. Note also that residential broadband traffic dominates dial traffic.

Klein: We are the blind men feeling the elephant.

It would be interesting to categorize the

flows. Always on Internet draws a lot of traffic during the night and day from Asia and Europe. Our low point is 50% of peak. For our network the low is at 10:30 GMT or peak at 02:30 GMT. During the summer the peaks are flat mounds, during the school year they are more pointed peeks. We start to get busy around 18:00 GMT. Having major franchise on the East and West coast does tend to spread traffic peaks out over four hour period. Heaviest use seems to come after prime time.

I suspect some of our summer traffic moves over to Universities during the school year. I would speculate that DSL would absorb more business traffic and probably exhibit a combined business and home pattern. Having done business Internet I can tell you that a few cable modems can suck down more bandwidth than the average office on a T1.

Kazaa et al eat a lot more bandwidth than any business application. As kc alluded to in http://www.caida.org/outreach/presentations/Myths2002/index.html. P2P is a killer app. and darn hard to stop because it morphs around filters and intellectual property laws. P2P keeps the cable modems humming during the day.

The big MSOs (i.e. TW/AOL, Comcast, Charter, Cox, Adelphia) don't do a lot of traffic at exchanges so they may be invisible. Until the demise of @home, most MSO traffic flowed over ATT. The number two provider for MSOs was probably Sprint. So the 30% of my MSO traffic probably never touched peer points but flowed via tier 1 private cross connects.

The RBOCs tend to buy interlata service from tier 1s. Therefore my MSO to DSL traffic is mostly invisible at peer points ... at least until recently. So far Shaw and Adelphia seem to be at the front of a wave of MSOs building networks and establishing public peering. SBC has become an aggressive peering entity. So if the government, academic, DSL, Cable and hosting networks peer around the large providers, how much do they end up controlling? Your guess? I have no idea what the elephant is like but I would be happy to talk with others about what I feel.

Nowlin: I agree with Joe [**Klein**]. Cable folks have taken a serious chunk of traffic away from the 'tier 1'club as dial-up users are canceling the second phone line in the house to save a few bucks. Measurement tools in use at SBC confirmed my gut feeling about this trend. We actively pursue cable networks as peers (hi Joe!) because they are typically behind one of the tier 1 club members and are growing much faster than the traditional transit suppliers. Telecommuters often use cable access if DSL is not available (I'm on RCN all day long) so it isn't just the nights/weekends that add to their traffic now.

Another reason why cable peers are good to pursue is that some tier 1 club members still enforce traffic ratios. Take the content out of their path. Many of them are in tough financial positions and are not adding capacity where they should. Some of the measurement tools available show they are loosing steam as customers do not renew long term agreements with such unstable vendors.

Most on this list are familiar with why a 'tier 1'wouldn't peer with SBC (multiple ASNs) even when we are able to meet them at more than a dozen common foot prints so forgive this backgrounder here for the rest. When I started at SBC nearly a year ago much of what exists behind AS7132 today was in a variety of ASNs for regulatory reasons and only seen via transit. SBC uses Sprint to carry IP traffic where LATA boundaries are crossed in states where local market-opening requirements of the Telecom Act's Section.271 have not been met.

Just because we are able to peer with a network, you should not assume nationwide routes can be exchanged. Legally they cannot be yet. All but 7 states have that approval now but California & Illinois are two biggies still in the works. Aggressive integration of the former Southwestern Bell territories has pushed out peering to +50% of the traffic today and y'all can figure out who the hold outs are using public tools.

Remember that California is another 11 LATAs not included in the 50% marker. In the Bay Area we peer as AS5673 and in LA as AS5676. The Ameritech (5 states) region also falls into this category. The rules are complex if you are not familiar with US telco regulations. Needless to say this has hampered peering efforts at most RBOCs.

BellSouth (on UUNET) & Verizon (on Genuity & Qwest) are stepping up their pace to build into Equinix IBXs right now. We (3 of 4 RBOCs) do chat often but I would say that is more due to the desire to improve the paths for our customers than to plan joint strategies. Just this week (Sept 16-20) BellSouth confirmed build plans for Equinix-San Jose, Equinix-Dallas and Equinix-Ashburn. Qwest (on C&W for the former USWest 14 states), is out of step with the RBOC crowd largely because of their deployment lead given timing of the USWest acquisition and that only one of the Internet 'hubs' (Seattle) is in their restricted territory.

SBC, as many of you know, is active in 4 of 5 planned at present and 5 PAIXs, 3 MAEs, 2 PBNAPs and AADS. http://www.sbcbackbone.net/peering/ is the place where this info is mostly public. We are trying to maintain management of peering in the Network side of the house vs. Marketing so links from http://www.sbc.com were not planned. I can't say this state of management will last forever but for now we are as flexible as able.

Equinix-Chicago, when the circuit is delivered, will also have public and private peering. Public on the GigE and with fairly open peering policies. Private on the ICE panel and with FE/GigE/OC-3 & OC-12 options just like in use at Ashburn, San Jose, LA & Dallas today. DSL traffic is growing fast and the smart carriers are beginning to realize Sprint will be the only path unless they peer. We are also the supplier for many federal agencies, universities, etc. so the traffic is more balanced than some expect.

Most nationwide ISPs have established

peering with SBC if they can meet us at three locations. 'Nationwide'is the line drawn between Sales & Operations which is why the three locations guideline is in effect. There are some ISPs who ignore my requests to peer. Do not assume they are all in the 'tier 1'club because they aren't. Many are angry for reasons that it would be counter productive for me to elaborate. Some are angry with the IX side of the house over which we have no control. We, SBC Internet Services, are nothing more than a customer of the NAPs by law. Welcome to the Telco Act of 1996 and the artificial boundaries that drain resources and cause confusion for all. Anyway the excuses run in many directions. Peering decisions defy prediction and can be very arbitrary.

For 6+ months I've been doing weekly top 10 in/outbound, top 20, top 30, top 40 & top 50 checks using Adlex traffic tracker. The Tier 1 club should be separated into growing versus shrinking camps. Only four networks from that club remain in our top 10 checks. Sprint, AT&T, UUNET & Level 3. The rest who claim it fall anywhere in the range of 20-60 now depending on inbound or outbound flows." Some other sources are adamant that while this data may be true for the source's network, it is not universally reflective of what is from other networks that track the traffic of these seven.

Woodcock: Ren, how long ago did C&W fall out of the top ten, and how far down are you seeing them now?

Nowlin: Around the same time Genuity did, so probably not long after much of the Exodus network was dismantled and customers started to run in other directions this summer. They are in the top 20 in some of our regions, top 30 in others. Some of AOL's RoadRunner holdings are higher in rank than C&W... Of course AOL's AS1668 is always in the top 10.

Freedman: I agree with most of what Ren says, but I most agree with this. We have 3 years of trending data to show exactly the same thing traffic-wise. Does your source use Adlex to look at quality of flows to different ASs also? CW and GENUITY fell out of the top 10 by origin AS over a year ago at Akamai (which looks at eyeball density) - but CW may be back in due to EXODUS nowadays... (Though that is still in many cases a different AS, I think). Of course, GENU and CW have continued to fall in % of total (using the potential metric I talked about) - but not in gigabits/sec - by Akamai's measure since we started measuring almost 3 years ago.

Adlex and Other Tools

Woodcock: Do you all use Adlex to look at quality of flows to different ASs also?

Nowlin: We use 'Traffic Tracker' from Adlex across 5 different ASNs now so the views are different and the latency & packet loss knobs are set for the affected ASN. (The views from NYC are different than those seen in Richardson, TX). We also have the traps configured with our ops tools. I use our configuration at keynote.com to poke around when packet loss traps start flowing in for AT&T (more frequent now than 3 months ago) and Verio (daily). Adlex is very very good. They have made so many upgrades over the past year. Mark@adlex.com can tell you more. Their office is located in Marlboro, MA.

Freedman: I've never seen their stuff in action, but Sockeye has a similar box that reconstructs TCP flows to figure out which prefixes you get dropped packets/small window sizes/early FINs/etc on. And Akamai has kernel hacks to the TCP stack, so we get 8-10 billion records/day of such information. Therefore useful aggregation is important.

But as part of peering, most networks can barely get good data on 'where' without moving to look at 'how good', and Adlex-type solutions can add that dimension, which is certainly interesting stuff. (For example those whom you DO pay for transit can get a nasty-gram with data they can use to make things better, and if enough people did it, Keynote and other active-measurement companies would finally go away, diseased methodologies and all.)

Diaz: It would be interesting to note, weather this reduced percentage was due to their restrictive peering policies, or simply because new customers are being added to different backbones. These new customers (cable and DSL users) are also doing significant peer-to-peer networking. It's easy to see in traffic by the way in the eyeball-to-eyeball backbones.

It seems the window is open once again for a new round of Tier 1s, as the % of the internet traffic shifts to the cable and DSL companies, they will wield enough "influence" to crack the door open. They also seem to have a much more open peering policy "at this time" which is also influencing this traffic balance.

Freedman: I'd say it's been more gradual in terms of some of the legacy Tier 1s losing traffic - and ATT became much harder to get peering with about 2 years ago, though L3 has been OK (of the upn-comers). Sprint and UUNET remain the most difficult, though Sprint I hear is harder and has been for a bit.

Diaz: As the percentage of Internet routes controlled by the legacy Tier 1s drop, so do their ability to enforce their restrictive peering policies.

Freedman: I'd say 'traffic' rather than 'routes'- not necessarily a good correlation except in the aggregate.

Diaz: I would be interested to see someone comment on the Yahoo/SBC DSL deal. I think we all know Yahoo is building a new backbone and has opened up peering. Anyone peering with them could expect to be able to receive high quality connectivity to some of the new services they want to offer like video streaming. Did they lose their patience and decide to go the direct route? Does not seem like a bad move on the surface, I was looking for some more details.

Freedman: Well, Yahoo and MSFT are both active in the network biz - MSFT

has in fact hired many whom people on this list know.

Traffic

Klein: [September 19]: This is as good as anyplace to share a few observations based on my corner of the elephant. We are seeing about 1 gig of exchanged traffic per 100K of cable subscribers.

Using the data from http://www.internetnews.com/isp-news/article.php/1446801 that would mean the top 11 cable companies in the US generate 88 Gig in exchanged traffic. Anyone have DSL numbers?

Doncaster: It's a much smaller sample size, but I've got just over 1000 DSL subscribers and my 95th% is 20 megabits.

COOK Report: I am a little confused. This surely is 20 megabits per subscriber per month? That seems low and one gig per 100,000 cable? Joe says 'exchanged ' traffic which must mean traffic that goes where? To transit providers and not to peers? Apologies if I have missed something really obvious.

Doncaster: We're talking megabits or gigabits/sec, not total bits transferred in amonth.

Klein: By "exchanged" traffic I mean traffic that needs to go to an external AS that is not a customer i.e. transit + peering as apposed to "on net" traffic such as a customer in LA exchanging a file with a customer in Miami.

One gigabit of peak data flow based on five minute SNMPsamples. So this is not a cumulative throughput but a peak "nightly" figure based on inbound traffic exchanged with peers and transit providers. Inbound vs. Outbound is used because as an "eyeball" network we get twice as much as we send. The average throughput for a hundred thousand cable modems at peak utilization on any given night (usually 10:00 PM ET for out network) is one gigabit.

COOK Report: You say "Inbound vs. Outbound is used because as an "eyeball" network we get twice as much as we send. Sorry again I am being dense. Please just a little more detail for the finance rather than network types who will read this.

Klein: A consumer-customer based network tends to consume more data from external networks than it delivers. Cable modems traditionally have a speed bias on the download side rather than the upload side. The same is true with asymmetrical DSL. In a pure html world, with no services provided on the customers machine, as viewed from the customer; outbound traffic is mouse clicks, inbound traffic is web pages. When I was at NAP.NAT we provided service to a number of large Chicago based ISPs. We would see 20% of the traffic up from the customer as down to the customer in dialup heavy clients.

Peer 2 Peer (P2P) programs have moved the ratios. On some cable networks the ratio is approaching 1:1 due to the volume generated by P2P. If I peer to reduce cost and latency or to make my network more robust, then the primary concern is inbound traffic. I am also concerned with ratios and volume since some large providers include ratio and volume in the peering criteria. Best ratio for peering is 1:1.

Manipulating our policy on hosting and/or giving out static addresses is one way for us to engineer the traffic ratios. At peering forums a network heavy in consumer broadband is known as an "eyeball" network. Coupling of interests makes for the best peering. Low latency is important to online gamers and online game providers. Eyeballs seeks content, content wants eyeballs. P2P traffic likes other broadband providers.

Our transit bills are based on 95th percentile utilization of the greater of in or out. Since my inbound traffic is greater than outbound traffic to me the outbound cost me nothing. I get billed for inbound not outbound.

Backbone Quality

Diaz: You are assuming that all backbones are the same. [...]

Stuart: That is a fine point - the subtleties of different engineering requirements for different applications are lost on the folks who make purchasing decisions (who think that only cost matters).

Spenceley: Is differentiation really lost? Or have the marketing folks managed to make the poorest engineered backbone look and smell like the best? Before folks can evaluate the advantages of one backbone over another, the industry needs to move away from buzzword compliance and 5*9 service "assurances".

I'd love to be Cogent's marketing department, imagine the fun you could advertising yourself as the least over-engineered backbone.

"A gig for $20k ... what more do want?"

"Other backbones offer you 4*9's, our 4*9s are the ones that count ... $19999 for a gig"

Genuity, in my humble opinion, is the best put together backbone for customer support and operations, yet only recently have they started to use this as a marketing tool.

When did peering stop being a marketing tool? Since the ascendancy of Abovenet no-one has really marketed their 'extensive peering' as an advantage, were the marketing folks really the first against the wall, or did they just give up and focus on price ?

Its still pretty early, but it smells like differentiation is coming, I'd guess it will really hit when backbones start to under perform due to load and capex/opex constraints. Look at 702. How long before a backbone in Europe starts marketing that they do have decent inter-region interconnects?

Diaz: Marketing people can whisper peering. We used to have major issues at

Netrail with marketing/sales people putting peering info into contracts and mentioning PRIVATE peering. The customer would then run with that contract somewhere else and we'd have to run to fix relationships with enraged peers. Peering is big time NDAmaterial and peering bilateral agreements have extensive language about disclosure. Setting up a page and showing trace routes to different backbones web servers achieved the desired results without breaking NDAs.

Doncaster; According to Alex Rubenstein (NAC.net), Cogent has increased its price for Gig transit to $30Kper month. Even if it were still $20K I wouldn't recommend anyone buy it considering the state that their network is in. They thought they were being smart buying PSI and NetRail for their Tier-1 peering. They've gone and overloaded the peering links, and my guess is Cogent can't force the acquired peers to increase the peering link capacity.

Cogent claims to be a "next-generation" Ethernet-based carrier. In reality their metro networks are OC48 and their long-haul OC192. Cogent will fail to meet their debt covenants by Q2 2003. Unfortunately for their competitors, Cisco will probably forgive the debt and let them start with a clean slate.

Diaz: While I'm not disagreeing with you, Ethernet is usually transported encapsulated. So for example, 2 gigEs in an OC48, 8 GigEs in an OC192. It's the reality of the world of networking, that most transport is SONET frame based. Even in DWDM and switching, the framing is SONET.

Avoiding Backbones?

COOK Report: from a slightly different point of view - A LOT of the internet COULD be knit together via ISPs going to exchange points and peering? The question is how much? How to figure the extent to which this architecture could replace tier one back bones?

Woodcock: Gordon, this isn't an either/or issue... Both are completely necessary. Let's say you've got a small ISP in Washington, and a small ISPin Tokyo. They're *never* going to have an IX in common, unless one of them is completely insane. The traffic between them *will always* go across a network which is in both locations, and be paid for as transit by one end or both. That's what transit is there for. Because it's cheaper than peering *for reaching remote destinations*.

John **Brown**: Not necessarily.

The IX that's being build in NM will purchase a DS3 to LAand connect at several points their. Backhauling those routes to the IX in New Mexico. Thus the providers in NM will bearable to peer with those in the LAX facilities. You could call it "limited transit" but in reality the model that's being looked at here is more along the lines of a Co-Op.

Woodcock: Fine. It's cooperatively owned and operated transit. But transit is transit. You're paying to transit that extension and reach non-customers on the other side.

Brown: All the members pay an equal share of the DS3 pipe. Similar idea is planned for NYC and there is a link heading towards Mexico City.

Woodcock: And here's the point at which this scheme always breaks. Tragedy of the commons.

Brown: With the shared pipe someone in NM could peer with someone in Japan. I believe PowerTier is at the LAAP. I might have the name of the ISP wrong. But there is a .JP [Japanese] ISP on the LAAP. In any event, if the traffic between the Wash and JP providers is such that getting a point to point is cost affective, why not? Maybe trans-pac isn't as cheap as US trans-national, but its possible.

Spenceley: In my experience such co-ops generally don't work or don't last. It's almost always the free time and good nature of a couple of parties that makes the IX run. At some point that free patronage will disappear and the exchange will slowly fall apart. The other option is the good natured parties take a small portion of the fee's to make their time worthwhile, of course this then leads to the other members feel they are getting screwed. In short they generally work to a point.

The SIX in Seattle is a noteworthy exception, but it doesn't run long-haul circuits or pay for co-lo.

Woodcock: The SIX no-cost/no-fee model is actually a pretty widely-used one. It works because it has no costs.

National Fiber Backbone?

Retzer On September 16: Concerning Gordon's ideas for a national fiber backbone ... I think that the problem to developing future markets and the Internet is the last mile, not the backbone. Unfortunately, I believe that the current economic situation and FCC policies will likely conspire to delay the rollout of really "broadband" last mile connections. The FCC's seeming reluctance to turn down requests for regulatory relief make it seem likely that the ultimate victors will be the ILECs and the cable operators. However, the incumbents are also generally in poor shape financially and I think the capital markets are loathe to invest in any telecom company given the lack of any clear outcome. The CLECs may be dead men walking but they are still threat in the eyes of most investors.

Lindqvist: What I think we have seen over the past decades has been a sinus-curve. For x number of years you will have a very fast development of access-technologies (one period of modems, another of xDSL technology), between each x-year period, you will see a y-year period where the backbones trying to catch up with the bandwidth is pushing the backbone technologies. I personally think that we are currently leaving one of the "backbone-periods" and are moving into a "access-period". Backbone bandwidth is currently more or less sold at below production cost. Carriers need to

start filling this bandwidth. I think that one of the reason for the telecoms collapse was that there simply was no application to fill all the bandwidth.

Retzer: It seems likely to me that the result of this will prolong the telecom depression and quite possibly put the US behind Canada and Europe in the deployment of next generation capabilities. I think this is a tragic, avoidable situation especially for an estimated half million telecom people currently out of work in this country.

It seems to me that the solution to the current malaise, which would also generate tremendous economic benefits is a national open access solution to last mile broadband. DSL and DOCSIS are great interim technologies but I really think that the nation needs widespread access to Gigabit Ethernet with voice and video as well as data over IP. Current technologies being demonstrated in Canada and a few rural communities in the US can provide these services over fiber — in a manner that would allow customers to select from multiple competing service providers. The beauty of this approach is that it would largely bypass local legacy systems that are currently delaying the deployment of advanced capabilities. I never thought it made sense to have a telecom deregulation act that focused on thirty-year-old central office technology. In sports terms, that's running to where the ball is now rather than to where the ball is going. It makes far more sense to leapfrog to IP and Ethernet both technically and from a financial perspective (Ethernet is an order of magnitude cheaper than most alternatives).

Lindqvist: As I said in an earlier email, fiber in Stockholm have been cheaper than a E1 (T1) for many years. Building out services of two Mbps or shared Ethernet infrastructure to private houses and multi tenant buildings have been going on for at least two years and is now pretty widespread. Unfortunately, most of the companies have trouble with their financial models. One of the companies claimed that the connection should be cheaper per month than the cost of a monthly card on the city busses and subway. This doesn't scale very well, but all the other providers had to jump on it. What is interesting though is that although the cost have been extremely low, a lot fewer users than expected have signed up. I think that the reason is that few have seen the need for it. There has simply not been an application there for them. They could equally well get DSL or Cable-TV. Especially CATV companies have been very successful because they could bundle the data service with the TV services.

Odlyzko: This is the key point. Most people have not seen much need for broadband. We have known this for a long time, from controlled experiments like INDEX at UC Berkeley, as well as from penetration statistics. (As of year-end 2001, something like 80% of US households had either DSL and cable modems available to them, but fewer than 15% of those took it.)

On the other hand, demand for broadband is spreading. In a presentation deck, "The many paradoxes of broadband," available at <http://www.dtc.umn.edu/~odlyzko/talks/broadband-paradoxes2.ppt>, I have statistics showing that broadband is diffusing faster than cell phones were at a comparable stage (cell phones took 5 years to grow as much as broadband is doing in 3 years).

Brown: What I have seen in the last six months though is that finally new services are starting to arrive. One company is now offering VoIP services, where the idea is that as long as you have a IP phone and a permanent connection you will be able to use it. This, of course, has triggered most of the access providers to launch their own services, and block independent attempts. Anyone remembers that pattern from the same industry?

Woodcock: In so far as I've been able to see, Kurt is right on the money on this one. I've seen it over and over as we bring up new exchanges. You'll be in some country that has only satellite access, and all the ISPs will be competing to see who can sell the cheapest unlimited dial. Start an exchange, and suddenly they've got two orders of magnitude more product to sell, and lower costs. Then all their attention turns to the access problem, and how to deliver broadband to people, so people will start to use some of the newly found excess backbone and exchange capacity.

Retzer: I view open access as more than just an idealistic solution. It would also spur competition for services that would lead to more investment, more innovation and lower prices. I think open access should operate via local Internet exchanges to foster local interconnection, which would greatly improve end-to-end quality and enable more advanced applications like telecommuting. Envision a future where you could pick your ISP, video service(s), and phone service(s) on demand from your easy chair. I think you'd pay extra for the knowledge that competition was giving you the best technology at the best price as well as insurance should a service provider go belly up. I also think that most communities would gladly consider some sort of loan guarantee if not finance a system like this if they could guarantee open access.

It would be very easy to implement this approach nationally simply by having the federal government offer loan guarantees for half the cost for deploying GE or better to the home for any community that guaranteed (or financed) the other half, as long as the resulting system were operating in a fully open manner. I think the result of this approach would be an Internet gold rush that would put the 90s to shame but with a difference — in this case I think people really could make money and you also really could deregulate.

I don't think it is surprising that business demand is greater than residential broadband for three reasons: 1) broadband is only now really catching on and 2) due to the lack of end-to-end quality and 3) because of Intellectual Property issues. The end result isthere really are no compelling applications. I'm hopeful that telecommuting will take off with broadband but if you need interactive two-way video conferences at a decent qual-

ity today you really need to have everyone on the same net. Internet video and radio are also being slowed down due to the Intellectual Property issues. Time will fix the first problem — critical mass in the next year or two even with a recession. I'm not optimistic about the other two problems.

Klein: You all may want to look at the articles in the Monday NY Times business section. The Lisa Hook article has some tidbits on AOL's P2P strategy. Note column 3 on page C3 in my addition shows 60% of broadband users doing file sharing.

COOK Report: Just for the record I agree 100% with virtually everything Jere **Retzer** says above. But I maintain that the backbones are also a problem.

Economics and Policy Issues of Commercial Backbones

Backbone services are necessary and the backbones are in really bad shape. Therefore what to do? We have bandwidth activity...true not enough....but we do have activity at the edges. what to do about the middle? Do you remember what Roxane said?

Please read what follows and ponder. This is what I was thinking about when pondering Andrew's remark that the cost of a national backbone was about ten billion dollars, I wrote my query regarding a national federalized backbone service.

Roxane said: I have been of the belief that IP technology represents a paradox. It is both our pathway to the future and economic kryptonite. It is so good, it is unfundable. You can read about my thinking on isen.com, especially "The Paradox of the Perfect Network", (or something like that). Bottom line, IP is such a perfect commodity, it guarantees its suppliers a loss on operations. For its job, of communications, this makes it perfect, because it neither imposes any preparatory restrictions on communications nor does it run out of capacity. Thus, the maximum creativity at the edges means zero value-add for the middle. Think of it as preserving entropy or something. This also means that no one makes money in the middle. This means our future well being is dependent upon something that no one benefits from doing. It also means that all telcos on the planet go broke, taking their investors with them. This is where it gets messy. How do we get out of our legacy investments? Who pays for the next build-out? (Hint: it has to be social as the markets are not about to touch this one.) End Roxane quote.

Retzer: [Roxane said: Bottom line, IP is such a perfect commodity, it guarantees its suppliers a loss on operations.]

Not exactly. The problem is that the backbone nets got ahead of demand. Absent improved services or technology, over the long run the current backbone market begins to approximate perfect competition which means that the marginal profit of the last bit per second sold will trend to zero. This does not say that everyone loses money, just that capacity quits expanding when the added capacity loses money. However this is overly pessimistic because demand will continue to grow rapidly for the foreseeable future (if we fix the last mile bottleneck). Eventually, as demand growth slows again, such as like now inefficient or financially weaker competitors go out of business and the most efficient competitors get by on profit margins that correspond to the risk level of the business, which should over time be about equivalent to the old phone companies.

However, the above also had a huge caveat that I think is not valid — 'absent improved services or technology.'Suppose we had very good interconnection at the metropolitan layer. Backbones could then differentiate based upon services & quality, which would encourage innovation and generate profits.

Odlyzko: Apropos Gordon's recent comments (and those from Roxane Googin he sent along), just as Jere **Retzer**, I am very skeptical of the claim that "... IP is such a perfect commodity, it guarantees its suppliers a loss on operations." First of all, can anyone name "a perfect commodity" that "guarantees its suppliers a loss on operations"? Commodities can be quite profitable, as your gasoline supplier will tell you. The main problem with IP is that there was a period of totally irrational overinvestment in the long-haul part of the business, and now that overinvestment is being squeezed out, with very painful consequences.

It is true that the backbones are unlikely to ever amount to much in terms of revenues or profits. It is simply not that expensive to provide high bandwidth pipes. If you look at the finances of companies such as Level 3, the total network build-out was typically on the order of $10 billion.

Googin on September 22: I can best try to explain why I view the backbone market as unfundable and a part of the perfect commodity by talking about the DRAM market. I have been following the DRAM market carefully for a long time. I believe that observing its dynamics enables me to see what is coming in the carrier backbone market.

To reiterate then the market that best approximates the IP carrier market is the DRAM market. In this essential commodity, no producer on the planet is making money. Nor have they for a while. Not only do they not cover their considerable fixed costs (a decent fab now is $3B), they don't even cover their variable costs. It costs about $4 to make a 128Mb DRAM, and they have been mostly selling for $2 for over a year. With enough of this behavior, most DRAM makers globally eventually come to rely on government funding, after they take their shareholders, bondholders and bankers to the poorhouse. The problems are identical: high up front costs to produce a commodity of questionable price. Since the fixed costs for all players are identical and high, and the variable costs for all players are identical and low, the guy with the volume wins on a unit price basis. The tendency therefore is to lose "just a little bit" of money to get to that

pole position. Since everybody keeps trying that same strategy, it becomes a global capital game of Chicken. Everyone involved loses.

Just review the history of Korea's Hynix. Ultimately, it becomes political bailout time due to the high costs of exit, not just from the jobs at stake, but from the inability to admit all of the losses were for naught. The more this happens, the more an economy behaves like Japan's, caught in the potential energy well of low interest rates and no investment. That is why our interest rate cuts are not igniting our economy. Until we break this logjam, there is no Return on Investment. The disk drive market, which I followed for over a decade, and made money on for my clients, behaves a little better because variable costs dominate the equation. There, starving vendors just manage to survive, and we get great disk drives.

Odlyzko on Sept 23: I do not agree with Roxane. People have on occasions made a lot of money on DRAMs. Further, IPis not that good an approximation to DRAMs. After all, bandwidth has both temporal aspects as well as geographical ones that distinguish it from hard good commodities like DRAMs. (If I want a Web page now, it is not too interesting for me to be told I can get it tomorrow. Also, if I want a Web page from Miami, it does not help me to be offered cheap bandwidth to Tokyo.) As a very direct counterexample to Roxane's claims about IP, consider Frame Relay. It is profitable, yet from a very high level it is hard to distinguish it from IP.

(As far as users are concerned, both carry bits, the things that users want, and
do it quickly. Further, Frame is used primarily to carry IP traffic.) The reason that Frame is profitable is that we have not had the same crazy overinvestment in it that we had in IP. (Now this may not continue to be true, since low IP prices are attracting traffic away from Frame and pushing down Frame pricing as well, but at least at the moment Frame is in the black.) Thus there is not anything inherent about IP that makes it unprofitable, any more than there was anything about railroads in the 19th century that made them unprofitable. It's just that when investment exceeded the demand that could be generated, companies went bankrupt.

COOK Report: No matter whom you believe there is still a problem. Backbones are failing and more will fail.

Hussain; This is the prevailing view. A recent Wall St report [Needham] states that 17 of the 20 leading wireline service providers are currently candidates for bankruptcy.

COOK Report: The investment bankers and equity analysts who tout industry consolidation and economy of scale, as in bigger-is-better, seem to think that three large players is all any industry needs. Is this where we are headed? If so, can three companies survive in the commodity back bone market?

Hussain: Om Malik's article in Red Herring http://www.redherring.com/insider/2002/08/bell082902.html
expresses a view that the three RBOCs SBC, Verizon, and Bell South together with AT&T are likely to make it through. But even these companies have substantial challenges and all of them have most of their revenue coming from traditional voice which still represents over two thirds of all revenue for carriers.

As pointed out by Andrew **Odlyzko** earlier backbone long-haul networks are likely to be low profit businesses. But they can undoubtedly be run as successful businesses. However, in my view it is not likely that the global market will be able to sustain more than a small fraction of the companies we have today worldwide. So there certainly will be companies that can successfully operate but they are as likely to be companies created out bankruptcies as any of the companies we can identify today.

In the US the RBOCs are not positioned to take a strong role in data. They are overwhelmed with challenges in Local and Long Distance and have significant constraints on capital. Their data revenue is still way behind WCOM and AT&T and they are barely present in the large enterprise market segment where all the higher margin data revenue currently resides. I would suggest that we don't look for the RBOCs to rescue us from the difficulties faced by WorldCom, AT&T, Sprint, Level 3, Qwest, C&W, France Telecom, DT, etc. - they've provided little or no evidence so far that they have an interest to move in this direction.

COOK Report: Won't the industry be a lot more healthy if as much as possible is done to lessen reliance on backbones? Isn't that what we are talking about here?

Hussain: It's clear from this discussion and others that the role of "dominant" Internet backbones is in decline. The important revenue is not in transit anyway which is a commodity in the US and on some international routes but by no means everywhere. Those carriers that have business models based on serving wholesale IP transit to downstream service providers in my view either have to rapidly adjust to the new circumstances or face going out of business fast. Many seem to be recognizing the problem.

COOK Report: Of course you have pointed out that you can by no means completely eliminate reliance on backbones. I understand that.

Hussain: Right, but we don't need government intervention. Some backbone operators will remain even at a global level to keep competition in place. At a national level in the US more and more service providers are finding that they do not have to operate under the old hierarchical peering order. They are observing that networks like C&W significantly diminishing in importance for peering and that they can find alternatives to route to any destination. So long as the choice of transit providers is broad and the pricing is competitive - and we've certainly seen that it is from the discussion on this list service operators at the edge of the network can grow.

In my opinion, there is no reason why we cannot sustain sufficient carriers to continue to competitively deliver long-haul US and international IP transit. I just don't have much confidence that this will

be done out of the so-called oligopoly of large backbone networks we discussed previously. Rather, I think these are in decline with WorldCom in Chapter 11, and the rest in growing levels of trouble. Because I believe that the current dominance is under siege and will likely collapse doesn't mean that I think that other carriers or other companies cannot or will not be able to successfully run business based on long-haul data traffic. It may be taking longer to happen than once thought but voice business is steadily declining and data business is still growing. Traditional carriers face the choice of having to significantly restructure or not survive.

COOK Report: While as Andrew says the expense is at the edge of the network much more than in the middle you disagree with Roxane's commodity analysis. OK. But I think you do agree that if money can be made in the middle on backbones, it is very, very little.

Hussain: I agree with Andrew's statement which covers Roxane's view about IP being a perfect commodity.

COOK Report: Now we never thought that there should be three competing interstate highway systems - one built by GM, the other by Ford and the third by Chrysler. We funded a single interstate system and let anyone drive any make car on it. One that is open to unrestricted use like the Internet. When any one of the national fiber backbones for the foreseeable future could handle all the traffic demands placed on it, how are we going to maintain three - let alone the six to twelve we now have?

Hussain: I don't think this Interstate analogy helps. Communications companies need to be global in scope if not always in physical infrastructure. To do so means that they must interconnect to and operate with other networks. So the nationally managed US infrastructure you propose has to be available to all competitors on equal terms. And I would imagine all US players would anticipate that they well have access to comparable national infrastructure in Europe, Asia and elsewhere on similar terms.

I just think that these kinds of schemes have never really worked out. If you take a look at international cooperative arrangements such as Intelsat you will see that they have been economic drains which have had to be privatized with a business framework that gives a very low probability that they can be successfully taken public any time in the next couple of years if at all now that the outlook for telecom is irreversibly changed.

Is There a Backbone Problem?

COOK Report: Hey. I am just an ex Russian historian turned bit jockey. I don't have the answers. I admit it. But I do think there is a problem. What do you think? Is there a backbone problem? If so what would you do about it?

Hussain: Here's my top line summary:

1. Peering under the old paradigm is over. The so-called oligopoly of 'dominant'backbones is in decline. Those that are in bankruptcy like WorldCom may not re-emerge as smoothly from Ch 11 as some anticipate. Others not yet in bankruptcy are headed there.

2. None of the major players in the US is showing signs of responding specifically to the conditions threatening their viability as businesses. They seem headed either to the court for protection against their creditors or to government for protection against competition.

Ugly. Because, there is no more capital from investors and the chance of government intervention is also extremely remote. So businesses will actually have to go through the pain of clearing up this mess themselves. And it looks like its going to take another couple of years.

3. There is not in my view a long term backbone problem. Both the US and Europe have survived the shutdown of many large networks and we are likely to have to survive quite a few more. At the end we will probably end up with a separation of wholesale long-haul capacity from 'edge services' as you describe them. True the long-haul stuff wont make much profit but it can still be run profitably even though it has not been so operated by the troubled telecoms of today.

4.The current FCC policy is completely out of sync with the circumstances. The FCC's 6 point plan to address the telecom sectors difficulties definitely needs to go back to the kitchen. The industry needs to have government policy guidance, the FCC doesn't seem to be able to do this itself and industry lobby groups seem locked in traditional outlooks.

Retzer: Fascinating points made by Farooq.

My initial instincts on this subject were let the market decide and I'm still inclined to go with that. That said, we have several interesting and potentially troubling dynamics at work here:

Current peering arrangements are largely secret — poorly kept but still technically secret. Seems to me that in a small oligopoly of 3 players that might be ripe for abuse. Reminds one of the California energy trading market — potentially worse. Could the surviving backbones use these secret peering arrangements as a barrier to entry? Could they use them as a lever to extract concessions/deals or perhaps take over local service? Could they use them to restrict bandwidth and artificially inflate prices at the expense of quality service?

Hussain: I agree with your concerns above. The purpose of the confidentiality in establishing peering agreements must be to protect proprietary information of the parties, information impacting physical and other security of the network, etc. The situation at present is without industry-wide guidelines where some networks publish requirements and others don't. It is one where many use the confidentiality of mutual non-disclosure agreements prior to negotiation as a shield for denial of peering on really quite obtuse reasoning for which there is no recourse for the adversely effected

party.

As a community, Internet Service Providers argued against any kind of government involvement in peering. The position that the industry was best served by permitting bi-lateral and multi-lateral peering arrangements to be self-managed by the industry itself has in my view now reached a point where either the industry works to fix the problems itself or allows the government to intercede to create some ground rules and guidelines.

Sadly, in the US at least I don't think either the service provider community or the government is going to do anything. In other countries the combined efforts of the industry, government and other interested parties have been sometimes quite successful so its not like it hasn't been tried or has never happened.

Retzer: 2) What are the incentives, if any, to provide end-to-end quality of service given the current arrangements? Why should one backbone agree to guarantee service to another's traffic? Seems to me that this arrangement may continue to provide incentives for each backbone to attempt to overbuild the network end-to-end. That's what happened in 2000. In little Portland, Oregon for example we had 36 franchises tearing up the same streets in an attempt to market to the same large customers. This clearly was not workable from a business perspective and also contradicts the whole spirit and value proposition of the Internet. Without any method to provide end-to-end quality of service how would we ever deploy the next generation of converged applications? The failure to provide appropriate incentives for QOS could cause the development of new capabilities to stall.

Hussain: Yes, I agree. There should be ways to provide commercial incentive for providers to support higher standards of service. When there are no financial incentives as in traffic exchanged on a zero-settlement/sender keeps all the revenue basis one network may invest in services which another will not wish to support without the financial incentive to do so. There's really no reason why ISPs can't work out revenue sharing schemes between each other in this regard.

When investment capital flowed freely service providers felt it was better to build the network themselves and be wholly responsible for the service end to end although this was never going to scale. The other approach was to acquire networks regionally and internationally. Perhaps in the current circumstances with sources of capital frozen up the incentive to pay for transit services above best effort may have a better rationale for service providers.

Retzer: 3) The global backbone networks are the classic case of 'increasing returns to scale' if such a thing really exist. This was a notion posed by economists in the 90s that new technologies were such that the more you make the cheaper it becomes. This contradicts classical economic theory of decreasing returns — the more you make the more expensive it becomes. The concept, which seems to make sense in PCs and bandwidth is that increasing sales leads to improved technology, which reduces costs and improves performance. We can expect this to be the case for Internet transit for as long as we can project. One interesting thing about increasing returns to scale is what they do to the competitive market place. Theory has it you only wind up with one, or at most three providers.

Hussain: I wouldn't be surprised to see the global infrastructure fragment. In such a case we'd lots of competition and continuing capacity builds for the major routes and little or no competition for lesser route/destinations.

Retzer: 4) If we were to wind up with one, or even three backbones would what would be the incentives to continue investment? Could a monopoly or even a small oligopoly use this as a means to extract 'excess'profits if they can erect barriers to entry? The barriers might be the peering arrangements (see #1). What would be the market and political power of such an oligopoly?

Hussain: I envisage backbone operators as needing to operate under separate business models from their customers. Its hard to see how this structural separation works here but in the UK the regulator [OFTEL] requires BT to make the network available to all comers at the same pricing levels that obtain to other divisions of BT.

So if we are headed towards a single large network operator and many smaller ones + the wireless networks + the cable networks etc. in the US all operators should be able to buy capacity resources on the same terms. But we're far from that point I think. Moreover, I think it is likely that we should see an emergence of a few low-profit long haul networks that are separate from the companies of which they are currently part. I'm not sure about how this separation happens other than being market driven.

5) Seems to me that we need and would benefit from competition but that the current arrangements may not provide adequate competition over the long run. I come back to the notion of publicly funded and/or subsidized last mile connections with public open access exchange points in the metro areas as a means to both renew investment in the Internet, solve the last mile bottleneck, save any number of telecom companies — and provide a new, open, competitive playing field.

Hussain; I'm not sure I understand why you see the need for publicly funded and/or subsidized last mile connections. What's that matter with competitive access pricing the the local network as with UNE-P?

Retzer: Several reasons:

1) Even the ILECs, in our case Qwest are in bad straits financially and given the current opinion of telecom in the capital markets I doubt will be able to get financing for local infrastructure. Public subsidy/financing would stimulate an immediate turn around from the current telecom depression. It would also make it possible for many more of the current providers to survive and prosper. Without this sort of approach, we are destined for

a very long, drawn-out telecom depression. A last mile solution would stimulate the entire economy.

Odlyzko: Hmm, yes, but would this be the most effective way to stimulate the entire economy? Just to be contrarian (and again, don't take this as my definitive word on the subject), at the end of 2001, there were almost exactly 10 times as many cell phone subscribers in the US as broadband subscribers (128 M vs. 13 M), although monthly costs were comparable. If one wants to stimulate the economy, shouldn't one then subsidize cell phones, which seem so much more popular?

Right now the telecom sector is suffering from huge (and terribly misdirected) overinvestment. Whether the cure for that is in putting more money in, or waiting for the excesses to be purged, is the question.

Retzer: 2) The ILECs are petitioning hard to overturn the '96 requirement to wholesale network elements on the basis that the cable companies are not required to open their networks. If this happens, I think that a number of CLECs and local ISPs will be toast.

Odlyzko: Yes!

Retzer: 3) Neither DSL nor HFC are good long term solutions to "real" high speed connections. I think we should be going for at least GE with integrated voice and video over IP to stimulate next generation applications nationally and globally.

4) Even if both the cable company and the ILEC survive and get some money to upgrade last mile access, this leaves the potential for some anti-competitive deals with the eventual backbone oligopoly and the potential balkanization of the Internet. An open access solution at the metro level would enable more effective competition for services and transit.

5) I don't think consumers are particularly eager to shell out $80++ a month for integrated services from a single provider but that they would pay a premium for choice of several service providers, which they could get with an open access network.

Odlyzko: Interestingly enough, the average monthly residential bill for just the wireline phone service (local, long distance, call waiting, ...) is already around $70 in the US.

Stuart [referring to our interstate highway system analogy]: Ah, but the maintenance of the interstate highway system is done by the states. When you see road crews working on the interstate highway system, the people you see are employees of the state in which you are driving, and money given to the state to perform that maintenance was given to the state by the federal government with *lots* of strings attached - like legislating a speed limit, or setting a certain drinking age. No-one *makes* the states pass such legislation, but if they don't the federal dollars to maintain the highway disappear.

Likewise, no, you can't just drive any vehicle on it. The fine state police whose jurisdiction includes the interstate highways would be happy to cite you or perhaps impound your vehicle for failing to be street-legal (try taking an Indy car on an interstate - if you get as far as the interstate), or for failing to display the proper tax stickers (look downstream of a busy weigh station sometime, and see if you see a state police vehicle waiting for trucks who fail to stop in to be weighed).

I don't have an answer to your question, but the alternative you propose - a telecommunications backbone analog to the interstate highway system - would, in fact, replace the six to twelve we have now with *fifty-one* (fifty states plus the District of Columbia - heh), encumbered with so many rules and regulations that driving a station wagon full of magtapes would suddenly become an attractive option in addition to an amusing classroom example.

Diaz [referring again to our conjecture about one backbone]: You are assuming that all backbones are the same.

Stuart: That is a fine point - the subtleties of different engineering requirements for different applications are lost on the folks who make purchasing decisions (who think that only cost matters).

Diaz: Maybe that has been part of the problem lately, that there is no differentiator of late for any of the backbones. You have to allow that some players will build to different specs or needs, and the rest of us will see that as a draw. Cogent uses Ethernet, a different technology then others, and perhaps they are a better and much cheaper fit for the server backup providers. They may also engineer to different standards, so they are a cheaper cost per meg. That may work better for some business models.

On the other hand there are some companies wishes to do VoIP on a large scale, and they may be looking for highest quality IP, or ATM or whatever. They would flow towards a diff backbone. The issue now is that all IP is being treated like a commodity.

Stuart: It's being treated as if it were all equivalent; orange juice is a commodity, but there's room in the market for national versus local brands, value-added services (like calcium), premium services (like fresh-squeezed). "Commodity" isn't necessarily bad. The corner we've been painted into (or have painted ourselves into) is bad.

Diaz: Perhaps yahoo will change this, if they build to the end user so that they can bring enhanced services it might spur new growth or changes for others.

Stuart: I'm of the belief that a significant amount of the content that would drive broadband sales, backbone utilization, etc., is locked up in RIAA/MPAAstupidity. Delivering that content could be the next "killer app," but the RIAA/MPAA goons are so focused on getting every bit of revenue for themselves that the orifice remains tightly closed.

Diaz: I agree that if we could release more of the content out there, it would drive more backbone traffic. At the same time, some 'killer apps' would be created. If we can consider that video band-

width or peer to peer is a killer bandwidth app, then we could see backbone growth once again skyrocket.

Let's face it, I have a gigE built into my MACs now, what difference does that make when I'm stuck with a 786k DSL line, or a paltry 10meg in the office environment.

I think the reason why more technology like remote apps or remote desktops hasn't taken off is that it doesn't work wide area on these sized pipes. If we could increase the pipe size, then these services would take off — possibly driving additional growth.

Odlyzko: Yes, making things like music would increase demand for broadband. In the case of South Korea (whcih now has by far the greatest residential broadband penetration), interactive games and social factors, together with low prices, seemed to be crucial. (You might like to look at Izumi Aizu's paper "A Comparative Study of Broadband in Asia: Deployment and Policy," <http://www.anr.org/web/html/output/2002/broadbandasia522.htm>.)

On the other hand, you could not do everything at once. At today's prices for transit (say $150 per Mbps per month) shipping a DVD of about 4 GB will cost you about $1.50 in transit costs alone (and only if you run your connection flat out), so in practice you would have to charge something like $10 or $20 per DVD. (Peter Wayner's column in today's New York Times, "The packaging of video on demand," <http://www.nytimes.com/2002/09/23/technology/23NECO.html>, talks about this.

Retzer: Certainly, having DSLand cable modems on the last mile will help stimulate demand. However, I don't see latency and loss-sensitive applications such as videoconferences over IP really taking off until we can offer quality service end-to-end. So, for example when people say that they think telecommuting will be the killer app for broadband access if they have in mind using videoconferences with the office they are not likely to have the quality they need unless everyone is on the same network or unless they have some local/regional interconnection arrangement. That is one of the main reasons that we created our regional exchange point — to enable better quality between networks by circumventing a lot of router hops and miles. Since Portland is not a "Tier 1" city, we had no local interconnections so we started our own.

Stuart: I assume you mean "quality of service" in the form of prioritized queueing, bandwidth reservation, etc., across provider boundaries.

Retzer: Yes.

Stuart: How do you propose that an industry that can barely agree on terms to exchange packets with each other at all come to terms with either a uniform service delivery model to, say, all VoIP packets equally (or equally poorly, depending on your point of view), or to reserve bandwidth in each others' networks (the potential for DOS attacks there seems like it could be a wonder to behold).

Retzer: That is precisely the problem. They can't agree or develop a successful multi-tier business model. The best solution I know is to connect the nets together as directly as possible, flattening the structure and ensure lots of bandwidth at the interconnect point. This seems to work regionally. Fortunately, there seem to be some number of applications like telecommuting and tele-med that are more local than global.

Stuart: The notion of this being something you can get, end-to-end through arbitrary providers, runs counter to the point that David **Diaz** was making (to which I agreed) that some backbones are going to be built to optimize certain services at the expense of others, to hit a certain price point for a certain market. All backbones are not created equal - nor should they be, despite the insistence of some in the market to think that "commoditization" equals "equivalence."

I think there will be some point at which the services become "good enough" for widespread adoption without end-to-end QoS guarantees, bypassing the (intractable, I think) inter-provider boundary issue.

Lindqvist on September 20: On the topic of TE (Traffic Engineering)...I have long been of the view that operators that have to use this has simply sold a commodity (bandwidth in their own networks) at a to low price so they now have sold more of it that they can afford to produce, or buy elsewhere. I think that TE at an operator is a bad sign and you should stay away. Services are generally very cheap though.

I think that what we are seeing a split of the market into

(1) Really local ISPs. Covering a city or a country-side population or village. This is the operator that picks the traffic that is to small for the larger players.

(2) Regional players. This is typically the old European PTTs or very early (pre-'95) new entrants.

(3) Global "Telephants", Sprint, Qwest, AT&T etc.

Is this going to be the final state? Most likely not. The temptation to make new alliances like Unisource, Concert etc will be hard to ignore. Will such alliances succeed? I don't think so, but that is more due to cultural differences than to unsound business models.

A New Fiber Project in Holland

van Hulten: Today (Sept. 19) there was a seminar in the Netherlands by SurfNet entitled GigaMAN http://www.surfnet.nl/bijeenkomsten/gigaman/ (in Dutch, but the presentations have lots of pictures): bandwidth for local communities, by local communities. Numerous projects are being started today by collectives of users (research + education + health sector) who want to have access to a fibre infrastructure, but won't be served by the incumbents and new telco players, who till now are hesitant to offer Managed Dark Fibre (and want to sell their Managed Services). These community projects, not-for-profit right now, are putting in place a city-wide duct, fibre and GigE switch infrastructures. Participants can choose to put fibres into the system and they make a financial commitment to the project.

STOKAB (Stockholm example) was also mentioned as a model to follow, but not available in the Netherlands. In the Netherlands we may however have passed the point of no return since there are a lot of vested interests already from the networks that have been built in city centres in the past few years. The local communities cannot revoke the digging rights of the commercial networks, even if they would prefer to have a monopoly operator again. The city councils should perhaps not have sold out a few years ago.

In the discussion that followed the presentations, there was a heated debate between the commercial operators and the city of Leeuwarden. The project undertaken by the latter, a duct/fibre infra for not-for-profit organisations, is seen unfair government supported competition by the commercial companies. The reply of course is that these companies should not complain, they can also join in. But these initiatives will flourish especially if the commercial guys stick to their 'cherry picking'strategy (as they called it) and not connect more users. Cheap fibre around the city is a good thing: to build distributed exchanges and to grow traffic volumes. With more bandwidth available to the end-users, "Internet" seems to be a good driver for more traffic (whatever the app).

Woodcock: There was talk about the Leeuwarden project at the RIPE meeting which just concluded, and it was pretty roundly critiqued as being ill-conceived and poorly researched. People complaining that companies weren't building infrastructure in areas where there wasn't any measurable customer demand, and then coming up with ways to make sure that no company would ever come in to change that in the future. A bunch of folks who really fundamentally didn't understand either market economics or Internet traffic exchange, getting in a big muddle with both. It wasn't encouraging to see, particularly as the AMS-IX is _putting to a popular vote_ the decision of whether to commit economic suicide by joining in. Argh. The theory is that their management should be there in order to insulate the customers from such bad ideas, but if they can't identify a bad idea when they see it, I guess that's kind of a lost cause.

The meeting was really frustrating. Person after person got up to try to hand the AMS-IX guys a bit of clue, and they _just didn't get_ that they were completely abdicating their responsibility.

Lindqvist: Notice that what is being discussed in the Netherlands and what Stokab is doing is something completely different. Stokab has stayed away from selling anything than dark-fiber. They have on a few occasions floated the idea that they could also provide GigE or IP services, but that would then mean that we are back to the PTT days where a monopoly can cross subsidize services. Besides from being doubtful from both Swedish and EU law, they realized that they would then most likely lose their customer base as they would be direct competitors.

With the risk of starting a flame war, this is similar to the AMS-IX expansion plans. What is currently worrying me in Sweden is that the government is paying the national power-grid operator (who of course is owned by the government) to build a nation wide fiber network. In principle they are only allowed to go into communities where there are no commercial offerings of darkfibers. With the current fibernetwork that leaves around 110 communities. Current proposals from the government will also allow the power-grid operator to provide transmission (SDH) and IP services to these cities. This means that they will effectivly kill the market in these communities.

Here in lie the dangers. The Swedish government today directly or in-directly owns five networks (1. part in Telia. 2.Vattenfall - the government owned power producer owns the broadband operator Arrowhead that recently bought Song Networks a independent fiberoperator in Scandinavia. 3. Terracom is the government owned radio/TV distributor that has and is selling a lot of transmission capacity. 4.The national rail-road grid is selling bandwidth on it's network .5. The power-grid network). What I am afraid of is that these networks are growing their market share and we are currently seeing control over a large part of the bandwidth going back into the government after a successful de-regulation. This scares me.

Speculation on Proper or Improper Agreements for Peering and Transit

On August 31 to Sept.2, 2002 on the Cyber telecom list, **Miles Fidelman** wrote: There's all kinds of peering and transit agreements in place now, most of which are non-transparent. I expect that some of the meltdown in the telecom. business has resulted from the complexity and fuzzy accounting going on in these agreements (e.g., I accept traffic from you, you accept traffic from me, we both book it as income even though no money changes hands).

Sean Donelan: I don't know what has happened recently with peering/interconnection agreements; but through 1999 because so few lawyers or accountants understood peering, they generally required peering agreements to be treated extremely conservatively. Anything is possible, but because peering agreements have been high profile items for so long, I expect the peering agreements were kept "pristine." Instead I expect the aggressive accounting was kept in separate, "unrelated" transactions somewhere far, far away in a different section of the books from the peering agreements.

Fidelman: I know that some of the folks who were doing bandwidth trading got into trouble for double booking transactions (I sell to you, you sell to me, we both book the sale as income). I'm just guessing that some of this has shown up in peering agreements.

Donelan: Sigh, bandwidth trading is not peering. This is one of those fundamental issues which Enron et al never understood, and one of the reasons why Enron was almost completely unsuccessful at negotiating peering agreements with other networks. Bandwidth trading is based on the premise that bandwidth is a fungible asset. Peering is based on the premise that connectivity is not fungible. Over the years, companies liked to tell their funders they had acquired "peering" when in practice they had purchased some type of discounted bandwidth. Discounted bandwidth sales were ripe for abuse. On the other hand, peering agreements are designed to have zero net value to both companies. Peering sucks for companies like Enron which wanted to generate the appearance of "selling" stuff. The problem with "peers" is they don't generate "sales" so you can't "round-trip" anything. If you are generating "income" from a peering agreement, I suspect it wasn't a peering agreement.

Do ATM-based Exchange Points Make Sense Anymore?

Commentary by Bill Norton, Equinix Highlights

Editor's Note: Bill Norton has over the past half dozen years established a reputation as the guru of peering writing multiple studies and "playbooks." We invited him to participate in our discussion. He declined, we now surmise because he had one of his own underway. This 'essay' is a compilation of his three major comments on the NANOG list during August.

While from a technical point of view we arenot about to argue with any of his conclusions that peering through an ATM based mesh makes little economic sense the motivation on the part of Equinix to finish of the remains of ATM based MAE East and MAE West is clear. We also understand that some Bell Headed providers are happy to have the MAEs remain open. What ever the case, in any putative "mini-encyclopedia" information about his report clearly belongs.

Hi all -

I've been working with a number of ISPs on a research paper that builds on the previous peering research papers (Internet Service Providers and Peering, A Business Case for Peering, The Art of Peering, Interconnection Strategies for ISPs, etc.) that applies the Peering Modelling tools in a comparison of ATM and Ethernet-based Internet Exchanges. Both of these IXes are compared against each other and against the cost of buying transit. The paper applies recent price quotes for transport and transit, costs for ATM and Ethernet-based IX participation, to answer the question:

I'd like to speak with additional ISPPeering Coordinators and Network Architects (preferable ones that have experience with peering across both ATM and Ethenet-based IXes) to walk through this paper and help me check that I have the technical and business details right. I would need about 20 minutes or so on the phone to walk you through the paper, the financial models, the cost points, and get feedback on the conclusions...preferably sometime in the next couple weeks.

If you are a Peering Coordinator I think you will find at least a couple of findings in this research *very* interesting. In any case, if you can help, please send me an e-mail at wbn@equinix.com and let me know when we could chat.

Thanks -

Bill

PS - As with any these Peering White Papers, this white paper will be freely available once enough folks have walked through it and verify that we have things right.

Abstract

During the NSFNET transition from the Authorized Use Policy Internet to the Commercial Internet, several Network Access Points (NAPs) were created to facilitate the traffic exchange between the Internet Service Providers, two of which were ATM-based. Internet Service Providers were initially required to connect to three of the four NAPs in order to receive NSF funds (indirectly through their NSF-sponsored customers) during this transition period.

During the years that followed, this requirement was dropped and the costs models of Internet Operation have changed dramatically. Technologies such as Wave Division Multiplexing and Long Haul Fiber Improvements have led to radical a decrease in the cost of transport and a corresponding drop in the price of transit. At the same, the cost of peering at ATM-based exchange points has not substantially dropped in cost, leading to the question in the Peering Coordinator Community:

"Do ATM-based Internet Exchange Points make sense anymore?"

In this paper we apply the peering financial models to this question, using current market prices to compare the price of transit against the costs of peering at ATM-based NAPs and Ethernet-based Internet Exchange Points. We build upon the previous research on Peering by introducing the notion of an Effective Peering Range (EPR) to describe the "useful life" of an Internet Exchange. We also highlight a potentially costly EPR Gap, an interim range between Peering Capacity points where peering is more expensive than transit.

The financial models presented that produced the graphs are included in the Appendix so that ISPs can apply these cost models to their specific situation.

Editor- Norton commented again on August 15

I have walked about 30 people through the "Do ATM-based Internet Exchange Points make sense anymore?" white paper and have received some really good feedback, suggestions and price points to calibrate the Peering Financial Model. I have applied these calibrations and I am ready to release the paper for wider review, but I'd like to share first the assumptions and calibration points for the model along with a few of the more interesting observations.

The Business Case for Peering at an ATM-based Internet Exchange Point Peering looks pretty dismal in todays market. As I mentioned in an earlier message, the dominant issue is that transit and transport have dropped dramatically,while the cost of ATM-based peering has not dropped in kind. In todays market (from quotes shared with me) we see:

Assumptions and Calibration Points

Transit $125/Mbps with 500Mbps commit, $100/Mbps with 1000 Mbps commit.

Transport (DC-ASH) $2500/mo for OC-3, $5000/mo for OC-12

Eth-IX fees: $2500/mo for 1/2 rack and FastE, $5000 for 1/2 rack and GigE
Eth Framing Overhead: 6%

HDLC Overhead: 4%

ATM-IX fees: $11,000/mo for OC-3, and $26,000 for OC-12 transport and Port

ATM cell tax: %20

Effective Peering Bandwidth=75% average utilization of available bandwidth (this means we assume that ISPs (for policy reasons) upgrade the peering infrastructure when the average utilization is 75%)

These numbers are empirical and based on averages from the Internet Operations Community. The paper footnotes the sources.

Observations

When these numbers are plugged into the Peering Financial Models, we see that OC-3 ATM-based peering is "Effective" (less expensive than transit) for the very narrow range of 88Mbps-90Mbps. If an ISP can't send at least 88Mbps over the OC-3 to the ATM-IX, it would save money by simply buying transit. At 90 Mbps the OC-3 ATM must be upgraded. This narrow range leads me to believe that OC-3 ATM peering is simply not cost effective. Under the same assumptions (OC-3 into FastE IX), the Fast Ethernet-based Effective Peering Range is 40Mbps-70Mbps, a more reasonable range for medium scale peering.

Applying the model to the ATM-OC-12 we see the Peering Breakeven Point is 260Mbps; if you don't send at least 260Mbps to the peering population then you should prefer simply to purchase transit. This peering infrastructure scales to 375 Mbps at which time it must be upgraded. In this Effective Peering Range the cost of traffic exchange ranges from $100/Mbps down to $69/Mbps when the Effective Peering Bandwidth is fully utilized.

The same analysis applied to Gigabit Ethernet shows a much lower Peering Breakeven Point (100Mbps) with a broader range, scaling up to 448Mbps before the OC-12 must be upgraded, at which point the cost of peering traffic exchange is $22/Mbps.

The bottom line is that the cost of the ATM Peering infrastructure, and the dropping price of transit and transport, have conspired to destroy the value proposition of ATM-based Internet Exchanges. Ethernet-based IXes are less expensive and have a broader "useful life", defined in this paper as "Effective Peering Range."

As I walked folks through this paper I got the sense that most folks had not done this analysis and we opened some eyes here. Thanks to those who provided the empirical HDLC, ATM, and ethernet overhead figures. Including these provides a more fair comparison between ATM and Ethernet-based IXes.

If you would like a copy of the paper please send e-mail to wbn@equinix.com and I'd be glad to send you a copy. As always, I'd love to hear your feedback; that is how these papers become valuable resources for the community.

Again later on the 15th

As an aside, I especially liked this paper request:

"I'd like to see a copy of your paper - please fragment it into 48 byte chunks."

A couple points seem to come up from a bunch of folks:

1) Several folks said that they have seen transit prices at sub-$100/Mbps prices, some claiming the transit price quotes group around $75/Mbps.

While the lower transit price points do strengthen the paper's argument, I would point out:

a) there is a qualitative difference between transit providers,

b) from my conversations there were higher and lower quotes than my $125-$100/Mbps,

(A couple of people told me they were paying $350/Mbps, but they were at the tail end of a 3-year old contract that was signed when $350/Mbps was a great deal!)

c) terms vary and location varies (rural guys are out of luck with no price competition, and some markets like Dallas are still high),

d) I want to make sure that the reference transit price points in the Peering Model are representative of what is seen in the field.

The bottom line is that I'm pretty comfortable with these numbers; $125/Mbps seems to be a price point that people can accept as a reference point for the Peering Analysis. And I've included the spreadsheet in the Appendix so you can adjust the transit price points as you see fit.

2) I explicitly mentioned in the paper that I ignored the equipment costs, in particular the OC-x POS and ATM interface cards and the equipment that ISPs would place in the Ethernet-based IX. This was because of the difficulty in determining a reference configuration (Juniper/Cisco, what series, new or used?), the price (people shared that 30% is easy to get) for a reference platform and then the lease term or amortization schedule. Some said depreciate things over 18 months, most said 24-36 months was the norm. In the past I have punted on this equipment question, but enough people mentioned it as a hole in the analysis (and a benefit of the ATM peering model) that if possible I'd like to include it into the analysis.

So I guess I am asking for a base level reference configuration and price point that includes two router configurations for the peering model:
1) entry level router with an OC-3 card and FastE card to peer across an ethernet IX, and

2) next level router with an OC-12 card and GigE card to peer across a gigE IX

I would also need an OC-3 ATM and OC-12 ATM price point.

Round numbers are fine here as I'm looking for some reasonable number to plug in for equipment costs, knowing full well that everyones configuration will be different, and the spreadsheet will allow people to adjust the numbers to their situation.

3) Finally, several have pointed out that the decision about peering at an ATM fabric is not always a financial one. These were most common non-financial motivations I heard were:

-) Performance: "I need to peer with this ISP regardless of the cost of that peering traffic."

-) Contract Term: "We are in the middle of an n-year contract so we are stuck with the economics." (One ISP lost a peering session when the target ISP left, and is now left hanging in the wind with a fraction of their peering traffic to justify their peering. Moral: Before signing up with any IX, Make sure your target peers are not planning on moving out!)

-) Perception: "To be a 'player' you have to be at xxx-IX."

-) Let sleeping dogs lie: "If I ask my peer to change the peering session in any way, I fear they will use the opportunity to force us to re-qualify for peering."

Comments from Our Particpants

Diaz: I was one of the participants in the essay, and listed as Bellsouth. I have had links at the MAEs. I do believe that the MAEs are on the down swing, people are rushing out of there and the value proposition for new peers is small.

However, I find that ATM technology has it's place. It's a good technology, reliable, easy to understand, and there is extensive written material for newbies to learn from. Jeff Wabik of Ascend/Netstar fame was once widely quoted as having said."ATM is evil..." What he really said was ATM was evil for a particular scenario. Ascend having acquired Cascade, and ATM shop and 70% of Ascend's revenue, it created a problem in house for him.

ATM technology has a place. It allows that QoS of many types of traffic, which may not interest us in colos, but over undersea fiber links or at the enterprise level over T1s or fT3s, it certainly does. While I appreciate other technologies are in place to do this at layer3 today, ATM still works reliably and carrier class. The biggest problem has been support of the carrier class legacy equipment deployed in the field as companies fail and engineers left to start ups.

As far as the MAEs go, the ATM MAEs were a much-needed upgrade to the old FDDIs many of us had to deal with. I mean I still remember packet loss between FDDI switches 8 and 9 if memory servers me correctly! Problem was they were turned up too late in my view.

As far as the FDDIs, I believe they have some of the same issues that the Ethernet based exchanges have. UUNET complained about traffic being spoofed and jammed at their router by people they didn't have peering with. Broadcast technology can cause a problem. This problem didn't exist with ATM exchanges or Optical Exchanges. This was one of uunets quoted "reasons" for not upgrading their router and link at the FDDI exchange. You can draw your own conclusions. I was told directly in person. No technology is inherently "evil." It just may become outdated, or less cost effective or just plain bad in certain business or engineering scenarios.

Nowlin: When you say that "some Bell Headed providers are happy to have the MAEs remain open, " I can assure you SBC, BellSouth and Verizon are not interested in the long-term health of the MAEs - including MAE-Dallas. Perhaps 'legacy providers' may be a more accurate listing?

COOK Report: Perhaps <smile>.

Klein: Any exchange technology using components that are sold to a mass market can archive superior economics vs. those sold to a narrow market. Gigabit Ethernet is used in corporate LANs, campus backbones, and even some SOHO.

FDDI and ATM failed to penetrate the larger markets. Mass produced chip sets are what makes the economics possible. An old computer saying goes "sand is cheaper than iron".

TCP/IP can ride on the cheaper networking technologies that archive an "economy of scale" that switched TDM telephony technology will never be able to touch.

Cheap, reliable, widely deployed Gigabit Ethernet makes for cheap, reliable router interfaces, and cheap reliable switches. Even if the router builders charge by bandwidth (wich they tend to do), they can make higher margins on interfaces built with less expensive chip sets. A survey of ATM chipsets above OC12 speeds enforces this view.

Understand the "economy of scale" on the level of the interface chip sets and economics becomes clearer. Your facts are correct; yet the depth of convergence of self-interest in moving away from the MAEs has complex undercurrents that span a broad spectrum of economic, political, and engineering issues.

Diaz: [. . .]All I was saying was that

ATM isnt a bad technology, like all technology, it depends on the scenario. In some cases it is bad and in others it is still useful, although currently out of favor.

Movement away from the MAEs may be as simple as they have lost critical mass now. Key backbones have moved out and away from the MAEs and therefore new players will find the value proposition greatly diminished.

Freedman: It is still true for many tens of participants at MAE-East ATM though, that the dollar per megabit cost they get on the MAEs is effective enough that it blocks them from going to an Equinix exchange. I agree that this may change over time, but most that I know who have MAE-E and MAE-W ATM connections are not unhappy with the dollar return.

I've been disagreeing with Bill about this (with the math) for years now, though.

Spenceley: I'd tend to agree with Avi. There are still a number of people I want to switch bits with that require a MAE connection.

The MAE's have become somewhat unattractive due to WCOM policy of slugging participants with a zero mile loop of about the same cost as an OC3 port. For those who avoid this, the MAE's can still make economic sense.

Freedman: Worldcom can generally be talked out of this if you're forceful enough... One other note - a lot of people are troubled about the economics of going to Equinix because (as far as I know) it's still only possible to get cheap (<= $3-5k/mo) loops over to Equinix by using the class of metro ethernet/OC carriers who are all likely to go bankrupt in the next year.

Spencely: Agreed. The price per Mbps Bill uses generally doesn't factor this. So yes you only need XXMbps to make an Equinix exchange cost effective, but costs are not always a constant.

I would love to peer at IBX's and if I had to build a new network it would be the place I started, but there are few (if any) decent networks you can't reach with a couple of MAE's and a PAIX, for me they have the advantage that the economics have already been justified.

Are there really that many people pulling out of the MAE's ? Cogent has made noise, but I can't recall in recent times having lost a peer (aside from Exodus)

Freedman: L3 did, but few others that I know of. A ton of people pulled out of-Pennsauken but that was mostly because noone at Sprint would respond any more re: cross connects, which were at one time promised.

Doncaster: What exactly DOES it cost to connect to MAE-East in DC? AADS is at least open enough to list their (ridiculous) cost of ~$6K/mth for an OC3 connection on their web site. Telehouse is responsive enough to give out list pricing of ~$1500/mth for a 100M FE connection to their peering switch.

Freedman: Without breaking NDA, I can safely say it's about $7500/mo for an ATM MAE OC12 and double that for an OC12. So at 120mb for an OC3 max, rounding to a 100mb peak to be safe (allow for some burst without congesting everything), that's $75/mb - but assumes you can do non-CBR VCs, which most peers will do nowadays. So cost/mb on an OC12 is roughly half that if you can mostly fill it...

September 25, 2002

Lack of Broadband Infrastructure Now a Bottleneck Holding Back IT industry

13 Micron Technology as Part of a Hardware Revolution that Brings PC Economics to Telco Switching

Googin Sees Inter-related Tech Revolutions Under Way that Will Create the Real Time Corporation but Cannot reach Maturity Until Enough Broadband Infrastructure Is Available to Enable the Network to Function as a Computer Backplane Highlights

Editor's Note: We interviewed Roxane Googin in New York City on May 19, 2002.

Googin: Some may look at the cost of LEC infrastructure as a barrier standing in the way of broadband There are significant changes afoot above and beyond fiber that are going to knock the foundations out from under the price of LEC infrastructure.

Cost Reduction from New Semi Conductor Technology

One of the big cost reductions that we are going to get will come from semiconductor technology. In order to get to 10 gigabits you generally use silicon germanium or gallium arsinide or some other exotic material. The ASPs are high. But the semiconductor industry is now going from aluminum to copper. Copper will make things faster. It is also moving from .18 microns to .13 microns. At .13 microns you have changes in the dielectric layers between the metal runs.

You have an issue with capacitance. It is like a battery. When you have two plates and some material in between - even air- you have energy storage. It functions like a battery as it captures the electrons and stops them from moving. The power of a capacitor goes up as a function of the inverse of the square of the distance between the plates. As lines get closer, it goes up fast.

However, you have these puny little wires that are only angstroms wide. You can't even get many electrons through them. As you make things smaller you increase capacitance. To counteract that growing force, the material between the conductive plates must have a characteristic known as "low K" dielectri (a low "dielectric constant"). They are having to change the material that they are using to achieve the low K dielectric because capacitance is getting unwieldy as things grow ever smaller. With .13 microns this stuff just doesn't work well using older materials.

Once you change out the material at .13 mircons, and as you move to Copper wires, everything speeds up dramatically. Moving to the smaller size allows the speed of these things to sky rocket. Moving to .13 microns, to Copper, to low-K dielectrics and to 300mm wafers all at once is the biggest change the semiconductor industry has ever gone through in its 30-year history. This transition has been under way for a while, but the people who follow this don't follow telecom. Consequently they aren't making the connection between the two.

These changes will have a profound impact on the cost and performance of transport equipment and especially metro transport. AMCC had been using gallium arsenide and have switched to become 80 per cent CMOS. What they both have said very clearly is that if the speed of telecom transitions had kept up and we were moving on to 40 gigabits, you would need the gallium arsenide. But since the telecom industry has stopped in its tracks for a technology generation, that CMOS has now caught up. Now you can do 10 gigabit physical layer chips, framer chips and network processors in CMOS rather than in the more exotic gallium arsenide.

This is the same CMOS that Intel uses in PCs. Don't forget that Intel is also moving to .13 micron copper and so is AMD. The processor industry is actually doing OK in what is a tough transition. The industry has moved about 5% of its production capacity to .13 micron, Copper processes. The yields are still questionable. You have what are known as high defect densities on the chips. One problem is that the low k dielectrics are gooey. They are runny. You are using between 6 and 9 metalization layers and your geometries are getting very narrow. As you etch and polishm and then reapply and etch and polish and then again reapply, you get damage to some of the lower layers. This is what is holding back progress.

Meanwhile, Applied Materials (AMAT) dominates the industry accounting for 30% of the semiconductor manufacturing equipment business by revenue. If you look at AMAT's first quarter announcement, you will find that in the January 2002 quarter, they had one billion in revenues and their orders were 1.2 billion. This second quarter of 2002 their orders were 1.69 billion. The nearly 400 million-dollar increase is first sizable increase in their orders in nearly a year.

COOK Report: Does this represent a shift towards everything being manufactured with the new technology?

Googin: There are two things going on with their numbers. They are a little suspect, but they have gotten people's atten-

tion. After close to two years of sequentially down orders and revenues all of a sudden they are growing. Even in a year ago their revenues were something like 2.6 billion. So in the last year they have gone from 2.6 to 1.6. In that sense such a decline is not good. Still from the bottom of the preceding quarter they have had a nice revenue spike. Now a year ago, even though their revenues were at 2.6 for that quarter, their orders were only 1.36. Now this quarter's orders are 1.69. Thus they are up both sequentially and year-to-year. This is the first time this has happened in a long time.

The one thing that makes me suspect their numbers is that they said DRAM was 10% of revenues and 24% sequentially. DRAM makers use aluminum because it has high capacitance to keep memories well charged. This is the one industry that is not going to copper although it will eventually. It was a really unnatural jump in the numbers however. I think it has to do with the high NIKs falling apart. So you have this jump upward that has people's attention. But if you look under the numbers it is not a clean next generation jump. However if you do keep looking through it, it looks like there is some progress.

Right now about 5% of the world capacity is in .13 micron technology. Everyone knows that if they don't make .13 micron copper chips soon - particularly in logic - they will not survive. These chips are going to enable huge improvements - for example a cellphone that runs almost forever on a single AA battery. I don't think people appreciate how revolutionary this technology is. It's a positive development. We could use some positive news. The orders for the technology sound, at the moment, like they are still in the pilot stage. But it looks as though production will begin to ramp upward in the second half of this year. AMCC may begin to book revenue on the shipment of these hot chips in the fourth quarter. But my guess is that it won't happen quite so fast. By the way, just because they ship doesn't mean anyone is going to have any money to buy them. That is a whole other problem!

The changes will, nevertheless, be significant. At the Spring VON Conference a guy from SONUS was talking. He had a diagram of these humongous 5ESS switches. As long as a large room. Then he said an RBOC today with one and a half racks can replace this equipment. The next generation will require half a rack. Half of a 19 inch rack will replace rooms full of gear. That is the power of this technology.

Among other things this means that if our regulators don't ban competition, space in central offices should be at a surplus. Power consumption should no longer be an issue and you really should be able to manage these things remotely as opposed to having to come in armed with a screw driver all the time. Many of the LEC arguments about why there is no room in their facilities will no longer withstand scrutiny.

Taiwanese Foundaries Driving Down Chip Prices

Furthermore your new switches will be founded on PC economics. Components for your switches and your PCs will, for the first time, come off the same chip fabrication line. Applied Microcircuits get their chips from UMC which is a Taiwanese foundry. There are two large foundaries. You have the factories (Integrated Device Manufacturers IDMs as they call them) where the Intel Motorolas and IBMs make their own chips. Then you have the foundaries. Their percentage of the world output has increased from zero, 6 or 7 years ago when they first came on line, to something like 30% now. In two years the foundaries may be doing 80%. Foundaries are really chip making outsourcing operations. Instead of doing it yourself, you send the design to Taiwan and let them do it for you. The foundaries don't design chips. They simply make them. A few years ago I sat next to an executive from on of the foundaries on a plane. He told me how they figured every ingenious way to undercut US manufacturing operating expenses.

Before this point the Chinese had a reputation for bad quality and it was the Japanese who had high quality. No one ever thought that the Chinese could ever get their act together in such a way as to raise their yields. But they have. What they do is one minute make chips for AMD and the next minute for AMCC. Consequently they produce enormous volume with their equipment. They can get the pricing economies that come from such volumes.

For a 10 gigabit switch you have your physical layer chips. The "phys" from AMCC that have to be the drivers. Then you have your framers which are different chips. These put your packets into frames and do your addressing. Then they have their network processors that they think of in layers of intelligence. Not only by the beginning of next year can they do the individual sections in CMOS as opposed to fancy and expensive gallium arsenide, but then they will also start to integrate all of these into one chip. You might want to guess how many gates it would take to put an entire phy, and framer and network processor onto a single chip that would cost about $250.00. That is the sort of price you should be thinking about for these things. It goes from millions of dollars to $250. A switch on a chip. That is the power of this transition. It is driven by PC economics. The finished product should be about as expensive as a PC.

Each foundary is getting more expensive. The cost to get into any high tech business no matter what kind it is seems to be going up. Networks are expensive to get into. Also software is expensive to develop the first time. The price tag to build a new foundary now has escalated to upwards of $3 billion. But once you get the volume up, the unit prices drop dramatically. Furthermore, wafers are going from 8 inches to 12 inches in diameter. Now it costs the same to put each wafer through all the different layers of processes, but you get a great many more chips out of the same processes. Three hundred millimeter wafers then should lower the cost per unit by 30%. The switch to increasing reliance on foundaries is driving down the

cost of the new chips in several ways. The biggest deal is being able to manufacture with cheap CMOS rather expensive gallium arsenide or silicon germanium.

Look at the history of how prices have been lowered in the computer industry. Look at mainframes which were TTL (Transistor - transistor logic.) They were bi-polar with very small scale, almost discrete, integration. What happened with the mini-computer was that you moved to medium scale integration and from bi-polar onto CMOS. This was something that enabled the use of air-cooled devices. This allowed them to go into manufacturing and also permitted interactivity for the first time. All these developments came with the rise of minis. There is a school of thought that says all progress rests on progress in material sciences.

The PC was really just a mini computer on a chip. With the microprocessor you just put all the logic of the processor onto one chip. Look at what the economics of doing this did to the industry. The revolution that is taking place now is one of this order of magnitude. What is really scary is that when you look at the SS7 that Lucent was still able to sell two years ago, we are going from the main frame straight to the PC. You go from these monstrous, expensive, hard to manage, power consumptive machines to relatively speaking a few chips in a box.

Switches on Chips Obsolete Circuit Switching

In this context, circuit switching becomes a bad joke because no one is going to go back and retrofit a circuit switch architecture onto these new packet switches. No one is going to redesign the circuit switched architecture to run on this material. The cost would be prohibitive. Lucent and Nortel are the companies that made the old telco circuit switches. Lucent and Nortel are no longer even designing the chips. AMCC is doing that. AMCC is like the Intel of the telco switch business. They design the chips for the products that Lucent, Nortel and Cisco make and at this point their chip design is for IP. Period. End of discussion. You have only a few chip houses that do this now. All the new movement is going into new layering technology and into outsourcing to bring that new technology to market. Nortel and Lucent are becoming like Dell. They are becoming box integrators and are adding less and less intellectual property all the time.

So here's what it all does. The computer business has taken over the telecom business. The way the computer guys do business is very different from the way the telecom guys behave. They are much more aggressive and competitive. Given current conditions, the telecom guys don't have a chance. When my customers ask me well what are we going to do, I reply there is a smart guy named Bill St. Arnaud who thinks we will go asset based. And they say that the whole movement of computing right now is to outsource what you don't want to do. The problem is that people aren't going to want to backward integrate into managing their own networks. This is one of the points of resistance into which we are running.

COOK Report: Then one possible metamorphosis for a carrier is to be come an operator for network outsourcing?

Googin: Yes. That is one possibility. Outsourcing is taking off like wildfire. In the most recent quarter IBM and EDS's outsourcing backlog increased between 40 to 50%. The management of EDS called it a bubble before they caught themselves.

SONUS makes class 4 switches. Apparently they can't make the class five yet. But the class 4 has fewer instructions.

COOK Report: What about the software that makes the five and seven run as a barrier? Who's replicating that?

Googin: I think that there is some sort of barrier. But I think the reality is that the software inside the 5 and seven switches will not be replaced. It will go away. They intelligence has to leave the network. The code in those switches is all about the 'intelligent' network. You can give those switches 2000 commands. I believe that is the number. But at this point for what?

COOK Report: So these new and cheaper switches will be set up to run SIP and Voice over IP proxies?

Googin: Right. A hard part about this transition is that it is not just a rebuilding of the old stuff to run better, faster, cheaper. It is building something new and different. No one but the monopolists are going to miss SS7. The developers hate it. I still don't think the software quite works which is a really interesting thing to start focusing on.

Now LEC CAPEX is down 40% compared to a year ago. In 2002 first quarter Verizon sold five billion in new bonds. Every deal was over subscribed. The interest was treasury rate plus 1/2% which indicates the perception of low risk.

COOK Report: As long as they can do this they can postpone the end by doing a kind of refinancing.

When Refinancing No Longer Works

Googin: Indeed. This is what they have been furiously doing for the past year. But all of a sudden the music will stop and when it does it stops really fast. Something happens and people get scared. That's how the debt markets work. They are really nice until they won't talk to you. Look at what happened to Qwest. They had a problem with their commercial paper and suddenly the company was weeks away from bankruptcy. Now they can't raise any new money. These debt guys aren't real swift. They are nice until they lose money, then they are GONE.

The first one that Verizon did was a billion dollars of five year paper. And they had 10 billion worth of potential purchasers. There is such a huge amount of money looking for a home. Everyone in the world (mid May 2002) is still plunking their money in the US dollar. If the dollar starts to slide and people want to

repatriate their money we have big problems. Our interest rates will sky rocket in an attempt to get money that our debt demands.

COOK Report: And if interest rates go up the telecom crash accelerates?

Googin: Yes, to the extent they have variable debt. Commercial paper's average lifespan is like three days. Its risk is nuclear. After Qwest blew up, it came to light that Verizon had something like 33 billion outstanding debt in commercial paper. That is callable overnight. When Standard and Poors demotes your bonds to junk, you have to pay your commercial paper back the next day. WorldCom just did go to junk and now they have to pay their commercial paper. This is why their stock is taking a bath because no one sees how they are going to do it. Your stock goes to two bucks and you are effectively closed out of the equity markets. You liquidate. Now last month WorldCom's market valuation was about $6 billion. Problem is they have 30 billion in debt. You could get something for UUNET certainly but whatever is gotten will pay back the bondholders. When you see bonds trading at only 50 cents on the dollar, it means that the markets don't think the bonds will be repaid. One of my clients was shorting the stock at $1.50 a share. You can short it all day at that price because it is going to zero.

They way bankruptcy works is that everyone is all smilely. I put out a piece called "Smile and Dial til You File." You smile at everyone. You tell them everything I fine, and then you file." The next morning everyone shows up and your stock is already worthless because you filed the nigh before to avoid a run on the bank.

You should look for the new switches that are going to come out to do so late this year or early next. That is the time frame involved.

COOK Report: Who will purchase these new switches?

Googin: The enterprises may buy them. There was an interesting presentation at the VON conference from the university of Alabama. It is mature enough so that the bleeding edge guys are using it. What I am hearing is that they are starting to buy IP PBXs and the talk on Wall Street is that Lehman has a big IPPBX going in with Cisco and that they are extremely unhappy with it. How mature is it? I don't really know. I think that the smaller business will probably start taking the IP PBXs.

Despite the Copper Loop Modernization of Voice Architecture Begins in Celluar

What I think I am seeing is that we are getting the modernization of voice architecture. Most of the transit of voice done by cellular systems is done with IP already. I am pretty sure that cellular long distance is all done as voice over IP. You are seeing IP economics beginning to invade the cellular business. The traffic can be routed to the cell sites within a city and aggregated there. My guess is the cellular backbone is not circuit switched. All the stuff that they cannot easily do with landline voice they can do with cellular. My guess is that this is indeed what they are doing. It is clear that VoIP works over backbones and it is also clear that, the lower the bandwidth, the more problems you have with it.

At the University of Alabama they finally did pull out their PBX which was like pulling out a mainframe. All your voice terminals are connected to the PBX. Consequently, you have to change all your phones out. They way it is done now is a rip off. For, as the computer guys come in, you will have handsets that can work on any PBX. To the computer guys this interoperability is normal. To the phone guys it is like: wait a minute, I cannot control my account!

In learning SS7, once you got to the APIs, you find that the are all proprietary. At every step you took, you ran into all these barriers that made working with different gear about impossible.

COOK Report: So there is an acceleration in the numbers of economic problems found in all layers of just about every system?

Googin: Yes. I think so. You need to figure out whether we are at the top of the "S" curve. This is why I have been so interested in the peaking of the voice minutes. Everyone says that this is just a little dip. I say no its not. It is like we are beginning with this 's" curve in reverse because as this grows, this other one has to shrink. When it starts going down, it goes slowly at first but then it gets pretty vicious and I think we are getting close to this now,

Most people still don't believe that the local phone company can be by passed to the point where it can't pay its bonds. I believe this is inevitable whereas most people believe it is impossible. They just cannot see how anyone can get around the LEC last mile monopoly. My belief is that to get to the next step to a better communications system we have to get around it.

Everyone says there is simply no way we can by pass that last mile. But they are wrong. And here the University of Alabama finally bit the bullet and pulled out their old PBX. One of the problems about the old style PBXs is that changes have to be done by hand. To move someone's phone is very cumbersome. In the new ones this goes away. You simply fill in forms in a GUI interface. They kind of justified their move on that basis alone. But the other thing that they got was 90% reduction in their voice transit costs. Instead stead of paying the phone company for a bunch of channelized T-1s, they put an IP circuit into the back of the PBX. This is what starts to kill the LEC. The university voice traffic then went into its data network instead of into the PSTN at one tenth the cost. This is but one example of what will start to bleed the LEC with more severity with each passing month.

You may slow it down. But you won't stop it.

COOK Report: Where does software fit in?

Googin: It will be hugely important because it will answer the question of how much is outsourced. There is web services software. Near term until web services becomes mature it is unlikely to affect the need for bandwidth significantly. For web services to work really well they all have to interoperate.

COOK Report: Clay Shirkey thinks this is two years away.

Googin: That is fair.

COOK Report: But if the cost of bandwidth continues to fall, it will presumably motivate people to finish this stuff so that they can use it?

Just One Part of a Multipart Revolution

Googin: There are a lot of reasons why people are working on this. First of all corporate spending on IT will stop until this works. There are like six simultaneous equations that I talk about. Servers are going from multimillion dollar single system image servers to commodity "blade' servers. Instead of paying 1.5 million you will pay $50,000 for the same computing power. All these are totally dramatic changes. The second change is from the PC to the handheld. Third is Software becomes Web Services. Fourth: The local area network merges into the wide area network. Fifth are the semiconductor changes I have just outlined for you in detail and sixth the storage goes from server attached to network attached. This new paradigm will first get us to intracompany. That is to say all the corporate applications will finally talk to each other.

There will be an intra-company stage where people will say I control my standards and applications. And then once each company figures it out, the next stage will be getting it to work between the companies. This is when it starts to hit the network and when you need solutions for the bandwidth problems. I think traffic flows will be extremely high because you will treat the telecom network like it were part of you computer back plane. It will be like another interprocessor communication. That is the ultimate. You won't care where your storage is. You won't care where your different processors are.

Think 130 nm is Interesting? 90 nm is Just Around the Corner.

Some recent public information:
http://www.dialelectronics.com.au/articles/a7/0c00fea7.asp
http://australianit.news.com.au/articles/0,7204,4899222%5e15321%5e%5enbv%5e15306,00.html
http://www.globalsources.com/MAGAZINE/EC/0205W2/INTEL.HTM
http://www.anandtech.com/cpu/showdoc.html?i=1677&p=5

As for 130nm, below is a link to doc to download on .13 micron process technology re: Intel
http://www.intel.com/technology/itj/2002/volume06issue02/index.htm

INTEL TO ADD SIGE CAPABILITIES TO 90NM MANUFACTURING PROCESS

Intel is adding high-performance communications capabilities to its 90-nm manufacturing process. These capabilities include the use of high-speed silicon-germanium (SiGe) transistors and mixed- signal circuitry. Intel said the integration of mixed-signal technology into its 90-nm manufacturing would lead to single-chip, hand-held devices that offer cell-phone, wireless-data-network and the evolving "personal-area-network" services.

Applications could also include network infrastructure equipment. Intel will manufacture all of its 90-nm communications chips on 300-millimeter wafers, enabling high-volume production and a substantial reduction in manufacturing costs.
http://www.intel.com/pressroom/archive/releases/20020916net.htm
Cited in Converge! Network Digest, v9n176 Sept 17

Why Broadband? It's the Latency Stupid!

-- says Internet Architect David Reed Highlights

In a private discussion a participant said "Beware of bandwidth fetishism. The interesting part of networking, is connectivity, not bandwidth. We need to develop more interesting applications that customers want, and I assert the interesting consumer applications will be mostly connectivity-based, not bandwidth-based."

David P Reed then made a very very important contribution: It's not connectivity and it's not bandwidth. It's the latency, stupid. (I coined that phrase a few years ago).

To be clear, what I mean is that end-user (task) latency is what drives the decision to use a communications experience. "Connectivity" and "bandwidth" are special cases of reducing latency.

Bandwidth reduces latency by reducing the time the network takes to deliver all the bits needed to all the places where the task needs to get done.

Connectivity (i.e. always on-ness and ubiquity) reduces latency because it reduces the time it takes for the user to get his/her task "into the computer communications system" by reducing the distance they have to travel and the time waiting for connections to be made.

(And of course other attributes reduce latency, like moving functions to the user's computer where appropriate, rather than the fetish of client-server so that one can bill for the server; like providing "presence" indications when the task involves getting other people involved).

The right way to think about what we call "broadband" has nothing to do with "broadband" - it should be called something like "when you want it, where you want it, as quickly as you need it to be". But I prefer "it's the latency, stupid".

People don't buy bandwidth, and they don't buy connectivity. They buy reductions in latency, and every network technology invention I know of that has been an important hit has significantly reduced the user-level latency for an important user task.

This leads to the observation that thinking about the Internet as content only is far too limited - the value is in reducing the latency in accessing the content the user wants. So companies with very limited content owning the access paths does very little for the user, even if there is huge bandwidth to that limited content. AOL's "broadband" initiative is really a small deal in the scheme of things, from a latency point of view, unless the stuff they offer is relevant to users' real tasks and reduces the latency, rather than being defined by what content AOL happens to own... People magazine gets read in the bathroom because that's where it fits in people's lives. There is no meaningful reduction in task latency for its users obtained by putting it online. In contrast, being able to use Google to answer a homework question or a research question in less than a minute is a big win for always-on cable modems. That's enormously quicker than a library, even if the library is in your house.

Try using end-user task latency as your way of thinking about this. I tried it, and I've never regretted it. :-)

Editor: Thanks to David for permission to publish

ICANN's Season of Delusions: Attempt to Spin Court Defeat is Rebuffed in IETF List

Froomkin Drives Sims, Cohen and Cerf To Distraction

Summary of Frommkin's "Form and Substance" Paper and Froomkin and Malamud Reaction to .Org Decision Highlights

The clique that runs ICANN has lived for the past four years in its own little dream world of deception and outright lies. The result for anyone who pays attention has been the destruction of trust in any mechanism of internet governance on the part of most outside observers. On the inside of ICANN meanwhile we have the reality distortion field of Joe Sims, Stuart Lynn, Vint Cerf, Mike Roberts, Esther Dyson and the rest of the IBM, ISOC, WorldCom clique who established ICANN in the summer of 1998. We shall visit that distortion field in this essay. The view of the insiders has been compared by ICANN observer Brett Fausett to the paranoia of Captain Queeg in the Cain Mutiney Court Martial. Fausett concludes If the blistering, paranoid rhetoric coming from ICANN officials sounds familiar, then you're probably a movie buff. We've seen this picture before. We can only wonder whether the ending to the ICANN story will be the same as The Caine Mutiny's.
http://icann.blog.us/stories/2002/08/17/whoStoleIcannsStrawberries.html

The Origins of the ICANN Clique

On August 20 an observer on the BWG list commented: “Also did you notice how the second technical evaluation team was a set of information officers from educational institutions? That's Lynn's background. He cloned himself.

We responded: Of course he did.

This is part of the small CIO group that begat Internet 2 and is acting to continue and protect the US research and education community’s original stake in the Internet. Back in 1987 IBM and MCI were joint study partners under MERIT on the NSFnet Backbone Cooperative Agreement. Dave Farber sheparded and guided much of it along moving from one institution to another as the idea was developed, put out for bid, and the award made. Dave Farber nominated Steve Wolff to become the NSFnet Director. Mike Roberts from Educom (IBM funded) coordinated higher ed participation in the project. Erich Bloch IBM was director of NSF from approx 1986-1990. BTW people I trust say Erich Bloch was the last good director NSF ever had.

In 1991 Roberts on behalf of Educom, joined CNRI where Vint and Bob Kahn hung their hats to form ISOC along with Terena (European research network.) Blockzil played a major role there. Van Howelling was the boss of MERIT from 1985 - 1995. He came to the university of Michigan from Carnegie Mellon preceded by a pro IBM reputation. One of his first accomplishments at Michigan was to throw the Amdahl main frame out and make Michigan an IBM mainframe campus. I have seen transcripts from the Confer computer conferencing system they used detailing how this was done in 1985 -1986. A decade later Chetly Zarko an old nemesis of van Houwelling played by means of State of Michigan FOIA lawsuits a key role in bring this into the day light. During the exact same period of time 85-86 I saw how NJIT completely revamped its design of EIES 2.0 in hopes that IBM would add NJIT to its list of mainframe donateable campuses. I saw that first hand working as the EIES expert for CSC on its NJ OTIS contract. Meanwhile Vint went from CNRI back to MCI in 1994. Circa 1985 or 86 he had developed MCI mail for MCI.

In 1996-97 van Howelling formed UCAID to do Internet 2. He did it with the assistance of George Strawn at NSF and Educause’s Mike Roberts. NSF money for the R&E communitty was funneled through Internet 2 which is run primarily by university Chief Information Officers. Educause was/is the University CIO association supported heavily by IBM. Roberts went from Educause to ICANN in October 1998. Lynn was CIO at Cornell and then CIO at U Cal Berkely. (Cornell was one of the NSF funded Supercomputer Centers from 1985 or 6 through about 1994. Its supercompuetrs were IBM machines.

Assuming there is another ICANN president after Lynn goes, I’ll wager it will be an Internet 2 CIO. Never has such a small handful of people stayed together for such a long period of time managing to control so much. MCI's CTO who built the NSFnet backbone was pretty much ignored by Bert Roberts circa 1992 after McGowen died. The CTO went off in a huff and became an adviser to Qwest. Qwest them donated a fiber pair IRU to van Houwelling and internet 2. As a direct result NSF's $50 million investment in the VBNS MCI highspeed backbone service died on the vine between about 1998 and 2000.

Do you see now why IBM's stewardship of the GIP under John Patrick was no accident? Mike Nelson looked out for the interests of IBM and Al Weis on Capitol Hill working for Gore and Hollings on HPC and NREN from about 1987 to 1992 at which point he was guided in to OSTP. With the second Clinton administration Nelson was moved to the Office of Plans and Policy at FCC to be ready for a Gore administration. But Mike meanwhile went to work directly for IBM Washington lobby in 1997 or was it 98?

Joe Sims, ICANN's Attorney, Has a Temper Tantrum

Editor: What follows is an outburst by Canadian Jonathan Cohen an intellectual property attorney and ICANN Board Member. Cohen is a typical example of the closed minded crowd that has been shuttled onto a rubber-stamp Board by the ICANN Clique. Mary Hewitt is ICANN's PR flak. In the email that Cohen responds to below she has apparently distributed the New Architect Editorial with ICANN. For the original see http://www.interesting-people.org/archives/interesting-people/200208/msg00070.html

From: Mary Hewitt [mailto:hewitt@icann.org] Sent: Friday, August 16, 2002 2:33 PM To: hewitt@icann.org Subject: new architect - ICANN of worms

"ICANN of Worms The Internet governing body is short on answers and out of time" By Christopher Null New Architect September 2002

Even if you're a casual New Architect reader, you've likely noticed that a new mug is staring out at you from above these words. I'm pleased to inherit the leadership of the magazine from my esteemed colleague Amit Asaravala. And while I'm not big on introductions, a few words of prologue seem in order, if for no other reason than to convince you that this magazine is in good hands.

My background includes editorial stints at Smart Business and LAN Times magazines, and before that, I put in several years in software development and IT management. I have an MBA from The University of Texas at Austin (and was born and raised in Houston), but speak with no trace of a southern accent.

I've been steeped in the Internet since the only "browser war" was between you and your copy of Mosaic. On a dare, I launched the movie review Web site FilmCritic.com in 1995, and much to my astonishment, the site is still kicking today. From my catbird seat, I get to observe the Internet as it impacts everything‹business, entertainment, pop culture, and mainstream society. It's the perfect vantage point for leading New Architect into a new era, one in which technology is no longer a curiosity, but a vital part of any thriving enterprise.

I hope you continue to enjoy New Architect as it evolves. I encourage you to write me with your ideas, concerns, and suggestions, or just to say hello."

"If any Internet issue demands your immediate attention, it's the battle that's being waged over the future of the Internet Corporation for Assigned Names and Numbers (ICANN). Of course, you can be forgiven for ignoring the protracted skirmish‹it is epic in its scope and extremely complex. In case you haven't been following the news, here's a sampling of recent ICANN developments.

Karl Auerbach, an ICANN director, sued to inspect the corporate records of his own organization. His case is still pending.

Congress opened a bitter inquest regarding the group, demanding accountability and a definition of ICANN's actual responsibilities. Congress is even threatening not to renew the Memorandum of Understanding (MOU) that gives ICANN a license to operate.

Former chairwoman and once-ardent defender of ICANN Esther Dyson pronounced the current organization "nothing but juvenile" and "a real cesspool."

Critics claim the group has too much power. Directors claim it has no power at all. Supporters say the group is hampered by its attempts to appease too many stakeholders. Critics say it successfully caters to the needs of no one.

And of course, critics also say that the organization's semi-secret meetings breed distrust and a lack of accountability. Supporters moan that all the group does is meet and talk endlessly, never making decisions or putting its lengthy proposals into action. When the group does get down to business, we end up with new TLDs like ".aero," ".museum," and ".coop." If I ever visit a ".coop" (reserved for co-operatives) in my life, I'll be shocked‹though the poultry industry really needs to hop on chicken.coop.

Even the ".name" TLD, which was supposed to be limited to personal first-name.lastname.name URLs, already has become corrupted beyond belief. Users have registered thousands of bogus sites, from greenbay.packers.name to sharper.image.name to santa.santa.name.

But annoying TLD issues are almost beside the point. The big question is what's going to happen when the bloated, power-mad organization does something that can't be cleaned up so easily. ICANN has only been around since 1998, and for almost a quarter of that time, it's been mired in "reform." The current operating budget is about triple the estimate of the original MOU (those jaunts to Ghana don't come cheap!).

At a mere 7,000 words, the latest ICANN reform document proposes a blistering series of changes, full of non-voting liaisons and advisory committees. Not surprisingly, Internet discussion has now centered on whether to scrap the whole thing and start from scratch. At this point, it isn't such a bad idea. Why not socialize the ICANN experiment? As loathsome as it sounds, even the IRS runs better than this."

From: Jonathan Cohen [mailto:jcohen@shapirocohen.com]
Sent: Friday, August 16, 2002 11:58 AM
To: hewitt@icann.org Subject:
RE: new architect - ICANN of worms

What a biased, superficial, distorted out of context, sensationalist piece of crap. The STAR needs this man to write copy for "Space Alien impregnates Pop Singer Madonna"..except even that may require some reasonable research and ..."Facts".. where did he get that MBA? and what does the B stand for?

Jonathan

PS please feel free to forward my com-

ments to Mr.Null and Void. Or the Board if they want a chuckle. God its a hoot being a Volunteer in Cyber-Hell!!

Editor: Still full of himself, Cohen wrote the DNSO General Assembly list at just after 3 PM on August 16.

"And will those who disagree stop villifying, name calling, cheap-shotting because they don't get their way. Will people from 'some' country stop invoking its world view on every aspect of ICANN and the Internet. Will those who don't like directions or decisions of ICANN stop whining to Congress or the DOC, will they take the time to check their facts? Bob Dylan said it in a song a long time ago ....."try spending a day in our shoes",as VOLUNTEERS, at considerable personal cost in time, energy and MONEY!! Try listening to the push and pull from every direction, the criticism, the Politics, the Rhetoric, name calling AND WORSE, while you try to do the Best Job you can. No one on the Board expects sympathy or "flowers"... But it would sure be refreshing to get some 'Balance' and some decent debate, where if you don't win you shake hands and try again later...But maybe that's "culturally biased" see you around the "Ranch"

Date: Fri, 16 Aug 2002 15:48:47 -0400 (EDT) From: **Michael Froomkin** - U.Miami School of Law <froomkin@law.miami.edu> To: Jonathan Cohen <jcohen@shapirocohen.com> [cc's were included in the original message]

I see you do have time to send email. Well then, allow me to check facts:

Is it true as alleged on the GAlist that the ITU was willing to pick up the charges needed to support the GAC, but ICANN chose to pay US$75,000 itself...in the absence of any need or contractual obligation to do so?

And why exactly should I be denied my right to petition congress and the executive for a redress of grievances? When it's the actions of the US government that empower ICANN? And it's ICANN which has chosen to disenfranchise me. I certainly have no vote on anything ICANN does, and you are planning to put the last nail in that coffin in October. I should sit quietly and take my medicine?

Incidentally, I find your comment that we should try to imagine what it is like to be in your shoes especially insulting, as it is your decisions which ensure that no one representing me will ever get to do exactly that. And of course, there's no 'try again later' once you've been shut out completely of the voting power. In fact, we had about 3 tries on getting user representation, all rejected by the ICANN board. That's a lot of tries already.

My cultural bias is in favor of representative structures and strict accountability. What's yours?

Editor: The next morning August 17, as part of the same mail list conversation, Sims wrote to Froomkin and cc'ed the mail list of the Domain Name Support Organization General Assembly. Now it is easy to imagine that Joe Sims simply snapped given how Michael Froomkin, in his writings at least, has done probably more than any other person to focus the thinking of those of us who take the time to look at how ICANN really operates as opposed to what ICANN claims to do. While we can see ICANN's manipulations, Froomkin has through his written and legal analysis of what the ICANN clique has done has written the Handbook that explains with great clarity why we ignore these people at our peril. He has made Joe Sim's prevarications much harder for Sims to get away with. Joe is mad and he's not going to take it anymore. Read on:

"Your "cultural bias," as best I can tell, is to try to get your 15 minutes of fame by becoming the Cassandra of ICANN, apparently hoping to rescue an otherwise unimpressive career by finding a niche where you can be perceived as the expert. And I have to admit that you have been pretty successful at that goal -- at least in the sense that you do get lots of attention, and invitations to testify, and calls from reporters. Is it exciting? Of course, there is one little thing: it does not appear that your constant criticism has much impact, on anything. Maybe that is because there are lots of other people, many of whom are also critical of some aspects of ICANN, who instead of just posting rants on the GA list and exercising their ego to run a webpage, actually dig in and work at trying to make ICANN more effective; those people do make progress in trying to shape ICANN more to their liking, because they invest the elbow grease it takes to get that done. Of course, you sneer at these folks as "collaborators," but the real sneers here from serious people are reserved for folks like you, who have no skin in the game, no willingness to invest even a little positive effort, and who insist that their positions are written on tablets and not subject to compromise. Junk like the stuff you post is exactly the reason why more people do not participate in this or other similar forums, because it is a waste of time. Of course, perhaps the worst thing from your point would be to actually succeed, since there would no longer be an excuse to pontificate on such weighty issues. An advance warning: I am not going to engage in a continuing debate with you or your fellow travelers on this point, so fire away all you want; you won't get a response. But your attack on Jon, who actually went to the (apparently useless) trouble to try to engage in a dialogue, has left me with an irresistible compulsion to point out that, despite all the noise generated by Froomkin Inc., there is no "there" there. Have a good day."

Joe Sims
Jones Day Reavis & Pogue
51 Louisiana Avenue NW
Washington, D.C. 20001
Direct Phone: 1.202.879.3863
Direct Fax: 1.202.626.1747
Mobile Phone: 1.703.629.3963

What Else One Might Wonder Caused Sim's Outburst?

We suspect that the venom that ICANN's highly paid corporate attorney

spewed forth had been simmering for a little more than two weeks. Sim's strategy had been grievously slapped down in a Los Angeles court room on July 29.

In Karl Auerbach's lawsuit when both sides (ICANN and Karl) called for summary judgment in the spring of this year, both sides, in effect, said to the court that the facts were not in dispute and asked the court to render a judgment as to what should be required of ICANN under law as a California 'public benefit' corporation. On July 29 the judge ruled decisively on behalf of ICANN Director Auerbach and on August 1 distrust of ICANN flared for the first time in a major way on the IETF mail list. The apolitical engineers of the Internet, listening to Dave Crocker lie and Vint Cerf prevaricate, came out and said enough.

Here is what happened:

Crocker: [snip] 2. Note that Karl won the suit, but lost the war. He is now subject to the conditions that ICANN had originally wanted to apply for his access. The only difference is that now Karl is under a court order to conform to those rules.

Froomkin: This statement quoted above in which it is alleged that Karl Auerbach got only what he'd previously been offered, is flatly contradicted by the text of the judicial decision. It is, quite simply, utterly false. In fact, almost the reverse is true: what ICANN got is almost exactly what Karl offered them months ago (the only differences being he has to give 10 days notice to disclose instead of 7, and some documents must be inspected on the premises). The judge stated that in several material respects, what ICANN sought violated California law.

See the text of the decision for yourself at http://cryptome.org/auerbach-icann.htm

See also http://www.icannwatch.org/article.php?sidà3 for a subsequent development.

One does not have to be a lawyer to understand what the following means: the court said ICANN's position "violates both section 6334 and Bylaws Article V, Section 21 because it deprives Auerbach of the inspection rights he has under law and imposes such unreasonable requirements as having to sign a confidentiality agreement and having to pursue burdensome review in any effort to enforce his inspection rights".

A few choice quotes from the court decision (note, Respondent == ICANN and the "Inspection Procedures" were ICANN's attempt to restrict a director's access to info):

[A] "Respondent contends that inspection rights of directors may be restricted by corporate inspection procedures and cites Chantiles v. Lake Forest II Master Homeowners Ass'n (1995) 37 Cal.Ap.4th and Havlicek v. Coast-to-Coast Analytical Services, Inc. (1995) 39 Cal.App.4th 1844 in support of its position.

Neither section 6334 nor section 6336(a) provide for or permit a corporation to impose restrictive conditions on directors' inspection rights and burdensome review when such rights are denied."

[B]

"Having considered the applicable law and the undisputed facts presented herein, the court concludes that paragraphs 3, 5, and 6 of the Inspection Procedures conflict with section 6334 and Art. V, §21 of the Bylaws by unreasonably restricting directors' access to corporate records and depriving directors of inspection rights afforded them by law.

Furthermore, Lynn's 10/5/01 letter violates both section 6334 and Bylaws Article V, Section 21 because it deprives Auerbach of the inspection rights he has under law and imposes such unreasonable requirements as having to sign a confidentiality agreement and having to pursue burdensome review in any effort to enforce his inspection rights.

Additionally, the Inspection Procedures here apparently have not even been adopted by the ICANN Board of Directors, but were promulgated by an ad hoc group of functionaries consisting of the Audit Committee, Louis Touton, Diane Schroeder, and Lynn (Auerbach Dec. Ex. 17, 18, 21).

Based on the undisputed facts, there is no triable issue as to any material fact and Petitioner Auerbach is entitled to judgment as a matter of law granting his Petition for Writ of Mandate. "

Auerbach: The writer of the above paragraph [Karl refers to Dave Crocker's assertion several paragraphs above] has his facts completely wrong. He is doing nothing more than parroting back ICANN's quite misleading press release. ICANN lost utterly and totally - not a single issue advocated by ICANN was adopted by the court. The court granted my motion for summary judgment.

Cerf: Both parties requested summary judgment

Auerbach: ICANN's so-called "procedures" and the "restrictions" issued under those procedures were rejected by the court with a kick that would be the envy of a world cup champion. ICANN is trying to spin its complete loss into some sort of Pollyanna story about how the result is what ICANN wanted all long. If this is really what ICANN wanted, then ICANN ought to have acquiesced to my offer to them back in September of last year.

Cerf: No, ICANN would have preferred to stick with the procedures that were established. However it seems appropriate to point out that the principal difference between the procedures adopted by ICANN's audit committee and the procedures recommended by the court differed primarily in the manner in which differences of opinion as to the releasability of material considered confidential by ICANN would be settled. ICANN recommended an internal procedure that, if not agreed by the director wishing to release them, would then be settled in court. Your proposal, that was recommended with one modification by the court, required ICANN to seek judicial remedy if given a 10 day warning, it disagreed with the director's proposal to release. In your original proposal you sug-

gested a 7 day period.

Randy Bush: So tell me, Vint. Maybe I am bit slow here. But, as ICANN is a shepherd of the public trust, other than personnel data, which are obviously confidential, why is ICANN not fiscally and procedurally transparent? Why did there need to be any of this pool-pah in the first place? Was there a critical shortage of controversy? [Editor's Note: Randy Bush is one of the most senior members of the IETF and early on was a hard core supporter of ICANN. However in the past year he has become quite critical. Still we have never heard him talk like this to Vint before.]

COOK Report: Vint, Dave Crocker got the rebuke from Karl and Michael Froomkin he so richly deserved. It's too bad you had to come to his defense.

Cerf: Karl's initial request was for unlimited access (fine) AND ability to release without limit any material he saw fit.

COOK Report: Vint you are mistaken. Please document you assertion. I have followed Karl's statements very carefully and he has been extremely careful to say that he never asserted such a right. Cite your source please.

Cerf: ICANN responded with a procedure to protect confidentiality.

COOK Report: It took ICANN 10 months to come up with a procedure that had Karl signed it would have been an effective gag order.

Cerf: Karl never took advantage of that (others did) but instead sued.

COOK Report: I understand there was one other a director named davidson who signed and then looked at the records, said golly gee i don't see anything here karl you just aren't a team player. well VINT no kidding karl sure isn't a team player..... he was elected on a platform that ICANN need to be cleaned up.

Cerf: I would note that his position as to access/release changed, at least as I understand it, after the suit was filed.

COOK Report: I am sure he will speak for himself but it is my understanding that his position on release of records NEVER changed. Karl gets non-confidential records that are in electronic form tomorrow Vint. Records ICANN considers 'confidential,' he will inspect on site.

And on August 2, **Froomkin** in response to Cerf: both parties requested summary judgment

Yes. And almost all of ICANN's was denied, and almost all of Karl's granted. Your point being?

[...]

Cerf: No, ICANN would have preferred to stick with the procedures that were established. However it seems appropriate to point out that the principal difference between the procedures adopted by ICANN's audit committee and the procedures recommended by the court differed primarily in the manner in which differences of opinion as to the releasability of material considered confidential by ICANN would be settled.

Froomkin: To the extent this is an accurate identification of the "principle" difference, it is also one where the difference between the two views is substantial, and ICANN had its head handed to it on a plate.

At INET, I tried to warn you this would happen. You clearly had been advised otherwise. That was very bad advice. You were so bought into it that you wouldn't even take the time to listen to *why* I thought what I thought. That was not prudent.

Cerf: ICANN recommended an internal procedure that, if not agreed by the director wishing to release them, would then be settled in court. Your proposal, that was recommended with one modification by the court, required ICANN to seek judicial remedy if given a 10 day warning, it disagreed with the director's proposal to release. In your original proposal you suggested a 7 day period.

Froomkin: This is what we in the legal profession would call a sop for ICANN. If your lawyers are telling you differently, I submit you are again badly advised.

Cerf: Karl, [the] restrictions are still there. If a director wishes to release information held to be confidential by ICANN, ICANN has a 10 day period in which to seek judicial review and restraint. At least that is the way I understand the order.

Froomkin: True. More importantly, a director has --and has at all relevant times had -- an independent, personal, duty to act in his best understanding of what's best for the entity. That is one which the law doesn't allow him to delegate. Or sign away, as ICANN proposed.

Cerf: Karl, even under the procedures adopted by ICANN to accommodate full review while protecting confidential information allowed you and any director full access to corporate records. The court did NOT validate unlimited ability of any director to unilaterally release confidential information.

Froomkin: Nor did Karl assert that nonexistent unlimited right to release anything. Rather he asserted his statutory right to exercise independent judgment, rather than having it illegally constrained by a procedure adopted in violation of both law and ICANN's by-laws. Doesn't that finding bother you? Illegal. Violation of by-laws. Doesn't it make you wonder whether other things you have been told by the same people who drafted this illegal document are within the by-laws might not be? Wouldn't it be prudent to worry about that now? Once bitten, twice shy?

The "unlimited right" you mention was never part of the case. Your lawyers tried to create this straw man. It was clearly falsified by the evidence, and the judge didn't waste any time on it. If this is what your lawyers told you the case is about, as opposed to it merely being a desperation litigation tactic ("blacken the

enemy"), which is what it looked like, then you have been very badly advised. I suggest you read Karl's pleadings. Or have an outside lawyer not affiliated with this proceeding, or with your current advisors' law firm, read them. And also the trial transcript. Then have a very serious and frank discussion with your professional advisors.

Cerf: In no case did ICANN seek to restrict actual access to documents but only to assure proper assessment of the releasability of anything considered confidential.

Froomkin: The law provides for this already in the duties that constrain directors. Karl went the extra mile before the suit was filed, offering 7 days notice before disclosure. The law does not require that (although it's a sensible and prudent thing for a director to offer under these circumstances). The court said, 'make it 10'.

Auerbach: Under the court's order ICANN must start delivering materials by tomorrow, August 2 - a mere 20 months after I first requested them.

Cerf: Some material is not deliverable but only viewable/copyable at the ICANN site.

Froomkin: I trust that ICANN will make every effort to be cooperative and to implement both the letter and the spirit of the court's order. It would be nice to hear you say that.

Another Froomkin Law Review Article:

Editor's Note: Michael Froomkin also published a citation to a preprint of a new law review article, FORM AND SUBSTANCE IN CYBERSPACE, on the IETF list. In this 30 page article from our point of view he does two important things. One he summarizes in crisp and clear language of about 500 words how the formation of ICANN was accomplished in a way that ought to outrage anyone who believes that governmental authority out to be accountable to the law and to those over whom it is exercised. Two he points out how even in a law review article Joe Sims and his co-author Cynthia Bauerly's "chief rhetorical tactics are obfuscation and confusion." In our opinion ICANN's foundations are so egregious that its only shield is "obfuscation and confusion." Keep your enemies off balance and it will be much more difficult for them to understand what you are really doing. For context the IETF exchange introducing Froomkin's "Form and Substance" paper follows.

On August 1 on the IETF list Joe Baptista wrote: "but failing that complete technical control of "." is in the hands of the Internet's end users. That's where the power in "." resides.

Froomkin: I translate this to mean that (1) we can point to any root we want (true, but not real relevant for Joe Clueless User) and/or (2) the root servers could choose to mirror something other than the DOC root (true in theory but not real likely in practice, cf. discussion of root server operators in http://personal.law.miami.edu/~froomkin/articles/formandsubstance.pdf

Editor: We republish below just over 2,000 of the paper's 16,500 words. We do so because we believe it important that our readers can see Froomkins analysis of the ICANN 'issue'and Sim's behavior.

FORM AND SUBSTANCE IN CYBERSPACE

by
Michael Froomkin

THE JOURNAL OF SMALL & EMERG - ING BUSINESS LAW [Vol. 6:

In this Response to the preceding article by Joe Sims and Cynthia Bauerly, A. Michael Froomkin defends his earlier critique of ICANN. This Response first summarizes the arguments in Wrong Turn In Cyberspace, which explained why ICANN lacks procedural and substantive legitimacy. This Response focuses on how the U.S. government continues to assert control over the domain name system, and how this control violates the APA, the nondelegation doctrine as articulated by the Supreme Court in Carter Coal, and public policy. Professor Froomkin then proposes that ICANN's role be more narrowly focused away from policy making towards true standard-making and technical coordination.

PLEASE NOTE THIS IS AN ONLINE PRE-PRINT DRAFT – ACTUAL PAGINATION MAY DIFFER IN THE PUBLISHED VERSION. [Editor: Complete preprint at http://personal.law.miami.edu/~froomkin/articles/formandsubstance.pdf

p. 106

In Wrong Turn In Cyberspace2 I had two basic goals, one descriptive, the other analytical and persuasive. The first goal was to describe how the Department of Commerce (DOC) employed a legal sleight-of-hand to achieve certain outcomes regarding the management of the Domain Name System (DNS), a key Internet resource. Thus, in Wrong Turn I carefully explained— and with the encouragement of the Duke Law Journal staff perhaps overfootnoted —the story of how the United States came to find itself controlling the root of the DNS, relied on by the overwhelming majority of Internet users. The Clinton administration, and particularly an inter-agency group headed by Senior Presidential Advisor Ira Magaziner, soon found itself faced with conflicting and irreconcilable demands. Internet people, such as Jon Postel, wanted to create a large number of new top level domains (TLDs). Assertive trademark and intellectual property interests—to whom the Administration was heavily beholden—strongly opposed this.

In its effort to escape this seeming impasse, Magaziner and the DOC achieved the paradoxical feat of keeping ultimate control over the DNS while maximizing the government's deniability and distance from the way in which the DNS was managed. In so doing, the DOC created a scheme in which it and its agent can make de facto rules that apply to all the United States (and most foreign3) participants in the DNS, despite the absence of

statutory authority from Congress. The result was an institution, the Internet Corporation for Assigned Names and Numbers (ICANN), that is increasingly able to impose its will on—that is, to regulate—domain name registrars, registries, and registrants, in ways that frequently benefit the trademark lobby and ICANN insiders.4

ICANN is, in form, a private non-profit California corporation and a U.S. government contractor. But the form of the U.S. government's relationship with ICANN is unusual, and the substance unique. The facts set out in Wrong Turn demonstrate that the U.S. government is the "but-for" cause of ICANN's existence, of ICANN's "recognition" by other relevant actors, of ICANN's ability to exact revenues from registrars and registries, and indeed of ICANN's continuing existence and relevance. In Wrong Turn I related each of these

p. 107

elements in perhaps tedious detail, including how ICANN and the U.S. government have entered into three different contracts. In these agreements the U.S. government lends ICANN power over the DNS, and ICANN provides what amounts to regulatory services for the government. Wrong Turn argued that these facts had, or should have, legal effect. Even though the form of ICANN's relationship with the United States was carefully crafted to disguise the fact, substantively, the DOC relies on ICANN to regulate those areas that the government fears or is unable to tread.

I also argued that, at least from a parochial, U.S.-centric, administrative law point of view, ICANN is a terrible precedent because it undermines the accountability we expect to accompany the use of public power. By vesting de facto regulatory power in a private body, the DOC insulates decisions about the DNS from the obligations (e.g., transparency and due process) and constraints (e.g., conflicts of interest, judicial review for procedural regularity, and reasonableness) that commonly apply to exercises of public power.5 Now that, thanks in large part to the energetic intervention of the U.S. government, ICANN has secured for itself a regular and contractually guaranteed income stream from the entities it regulates, it faces few external constraints on its behavior. Although firms that lobby ICANN as if it were a government body may face anti-trust liability,6 and ICANN theoretically might be seen as their co-conspirator, to date the chief source of external discipline on ICANN has been the looming possibility of U.S. government oversight combined with the background threat of the U.S. government exercising its right to take back all the powers and functions it previously bestowed on ICANN. 7

My second goal in Wrong Turn was to explore the legal theories that could—and, I argued, should—be used to right this departure from administrative regularity. The key conceptual move was to focus the legal argument on the government's role in DNS policy rather than on ICANN's actions.

p. 108

In Wrong Turn I began by arguing that so long as the DOC continues to control the root, the law—cognizant of the substance of the relationship rather than focusing on form only—requires the DOC to regulate the participants in the DNS via traditional APA processes rather than through contracts and winks. We know at least part of the story about the DOC's role in ICANN's formation because it is public. We know that at one point the DOC estimated that monitoring and assisting ICANN would require the half-time dedication of four or five full-time employees.8 Further, the DOC testified to Congress that ICANN "consults" with the DOC before its major decisions,9 and that in at least one case the DOC amended an ICANN decision.10 On the realpolitik side, ICANN very much wants to have full control of the root, and the U.S. government, after initially signaling that it would transfer full control to ICANN, increasingly waffled as to when if ever it would relinquish control.11 We also know that, while the government maintained this powerful club over ICANN's head, ICANN had in fact done pretty much what the U.S. government had said (in a formally non-binding statement of policy) that ICANN should do.12

These, and many other facts related in Wrong Turn, indicated that, despite a veneer of arms-length contracting, the DOC was either the instigator of, or the conduit for, ICANN's regulatory decisions, and as a result ICANN's actions should be fairly chargeable to the DOC. Thus, in Wrong Turn I concluded that the DOC's approval and acquiescence to ICANN's actions pursuant to the DOC's at least tacit instructions constitute regulatory actions that must conform with the APA.

Snip **p. 109**

If the nondelegation aspects of Carter Coal are still good law, then I think the DOC's relations with ICANN are that very rare case to which the doctrine applies. But, just as the nondelegation doctrine's history is bound up in the struggle over the New Deal, so today it is likely that any nondelegation argument is going to carry substantial political or constitutional-structural baggage. Thus, in Wrong Turn I made both legal and policy arguments. First, I suggested that, despite its seeming desuetude, the Carter Coal nondelegation doctrine had never been formally repudiated, and I cited modern state court decisions relying on it.19 Having established that the doctrine at least remains available, I then argued that ICANN's corporatist structure, its inbuilt self-dealing by design, and the regulatory nature of the services ICANN provides for the DOC, all combine to make it that rare and special case to which the Carter Coal nondelegation doctrine ought to apply. Given ICANN's recent behavior, that policy argument seems, if anything, stronger today,20 while the doctrinal picture remains unchanged.21

Snip - **p. 110**

FLAWS IN SIMS & BAUERLY'S CRITIQUE OF WRONG TURN

Although Wrong Turn focused on the le-

gality of the DOC's actions rather than ICANN's,22 Joe Sims and Cynthia L. Bauerly (S&B) seem to take the critique very personally.23 This is perhaps understandable as, more than anyone else, Joe Sims is responsible for the ICANN we have today. He launched ICANN even after his client, Jon Postel, tragically died, and through his advice and actions established its substance and style. He and his subordinates wrote ICANN's charter and by-laws, and then frequently revised the latter. He has

p. 111

presided at a number of ICANN meetings, and remains by some accounts its éminence gris. 24 Furthermore, since ICANN's inception on September 30, 1998, ICANN appears to have paid Joe Sims's law firm a total of at least $2,171,283.88 in legal fees.25 If one assumes an average fee of $300 per hour,26 that sum would work out to more than 7,200 lawyer hours, or well over three full-time lawyer-years; in fact, the number is likely to be substantially smaller because that figure must include recovery of hotels, international airfares, meals, and other expenses. But whatever the actual number of hours billed and hourly rates, ICANN now amounts to a substantial, valuable, and recession proof client.27 In addition to casting doubt on the wisdom of the course pursued to date, were the DOC forced to rely less on ICANN it likely would reduce

p. 112

ICANN's importance, and its need for expensive legal advice.28

S&B's chief rhetorical tactics are obfuscation and confusion. For example, S&B called my omission of most of the history of the Internet "misleading."29 Wrong Turn did not discuss the large majority of the history of the Internet because the history of the Internet at large is not terribly relevant to the issue of the DOC's legal obligations regarding its management of the DNS, just as the article left out the history of computation, of capitalism, and of the United States, all of which are also parts of the background. Yes, much of the Internet—the devices using TCP/IP and the programs running on those devices—was at all relevant times private. Yes, the World Wide Web was created and open sourced by Tim Berners-Lee. So? Many Internet services, the Web among them, are layers above the DNS. That has nothing to do with ICANN because ICANN does not (yet) have any functions relating to the World Wide Web. ICANN's jurisdiction thus far has been limited to the DNS and to IP numbering. Those functions, especially the regulatory functions, were, for many years prior to ICANN, performed by the U.S. government or by its contractors, primarily Jon Postel and his associates.

A related rhetorical device frequently used by S&B is the attack on the straw man. The problem begins in their subtitle, which takes aim at a contention ("ICANN . . . Violate[s] the APA") not found in Wrong Turn.30 I ask the reader to look in S&B for citations to Wrong Turn indicating where I supposedly said the things S&B put in my mouth. You will not find many of them, often because they do not exist.31 That is a serious failure in an academic article (especially one that claims to be a response to something), a literary form where pounding the table is considered a poor substitute for pounding facts and law.

I would stop here, were it not for the fear that someone, perhaps put off by the length of Wrong Turn, might read S&B alone and decide that the absence of

p. 113

a more detailed rebuttal was in some way to acquiesce to it. The following sections thus respond to some of the fundamental errors in S&B's response to Wrong Turn. S&B argue that ICANN's power is either low, or not derivative from the U.S. government. This claim fails to recognize the source and scope of ICANN's power—the U.S. government's recognition of ICANN plus the government's functional control over the authoritative root, a power based in part on the U.S. government's ability to veto any move by the root server operators. S&B's state actor argument is also flawed. ICANN, unlike an ordinary government contractor, performs regulatory functions for the DOC. The main issue is not, as S&B would have it, whether the APA applies directly to ICANN (it doesn't), but rather what constraints the APA and the Constitution put on the DOC's use of ICANN when ICANN performs public functions directly or advises the DOC on what actions to take. Finally, I touch on the nondelegation argument advanced in Wrong Turn. Even though S&B concede that ICANN makes policy decisions relating to the DNS, they concentrate on the Schechter Poultry branch of the nondelegation doctrine and thus fail to grapple with the doctrine's due-process strand, as articulated in Carter Coal. This doctrine imposes a fundamental structural constraint on the entire government's power to delegate public functions, like the power to regulate, to unsupervised private groups.

ICANN to Do Crony Re-delegation of .org?

Meanwhile we have another perfect example of giving “government's power to delegate public functions, like the power to regulate, to unsupervised private groups.” It looks like ISOC will be given the royal right to sell .org names. We quote in their entirety two brief articles.

http://www.icannwatch.org/article.php?sid=927

Old Internet Thinking RIP
Posted by michael on Friday, August 30 @ 11:13:13 MDT
Contributed by michael

Carl Malamud, one of the sponsors of the IMS proposal for .org, has posted his response to ICANN's tentative decision to give .org to an as yet non-existent body to be created somehow or other by its good friends at ISOC. The essay demonstrates why IMS's hopes, and the rest of us probably, are doomed.

See, the problem is that Malamud's entire essay is consumed with irrelevant Old Internet considerations like running code, technical merit, and whether it makes

sense to evaluate a program without ever looking at it. This IETF-style approach to the problem of finding reasonable solutions to problems has no place in the Brave New Internet of today where expensive consulting firms decide that proposals produced by expensive consulting firms have the most merit, where merit is defined as producing familiar-looking paper. Only a dinosaur would have failed to notice that "the ICANN .ORG review mechanism literally restates the ICANN new open gTLD contract award order.". Only an ostrich would fail to see that ICANN has learned nothing and forgotten nothing from the gTLD rollout debacle. Recall that mere factual errors were no reason to upset the gTLD allocations.

Read Malamud's essay. Don't miss the Grrrrrreat slides. Weep or gnash your teeth. There's not much else you can do now that the ICANN Board is preparing to undermine just about every form of outside accountability that might be brought to bear on it.

It is possible to argue that ISOC will do a perfectly competent job on .org. It just won't have the lowest prices. And, if Malamud is to be believed (I am not competent to judge, but he is), it won't have the best technology. And it certainly won't do anything to increase competition in the market for providers of registry services. But you can't have everything, can you? And no doubt we'll have the comfort of knowing that .org is in safe and familiar hands as soon as ISOC actually gets around to telling us who will be involved in running this new corporation they are planning to set up Real Soon Now.

I'd feel somewhat less bad about that argument if ICANN would release the data that might back it up. Unless of course the data show something else. Assuming it exists.

Carl Malamud's Take

My Fellow .Organisms: http://not.invisible.net/signals/bin/000270.shtml

Many of you have sent signals asking what we thought of the ICANN Preliminary Evaluation. Three committees ([1], [2], [3]) generated rating matrices which was summarized in a staff-generated meta-matrix ([4]) which yielded a single metric. While the IMS/ISC bid received extremely high marks in supporting areas such as vision, experience, innovation, service, and commitment, the technical evaluation rated us somewhere between totally clueless and moderately brain dead.

Rather than engage in a blow-by-blow evaluation of the evaluations, it seems more appropriate to pull up to 50,000 feet and explain why there is a fundamental difference of opinion about how to do what we call in the trade "technical due diligence." Since "a demonstrated ability to operate a registry of this scale" is the primary technical criterion, all the supplicants prepared appropriate paper trails to document such a capability. Much to our puzzlement, we haven't received a single email, phone call, fax, or chat room request to view source code, do a site visit, examine log files, or get an ssh login to look around our systems.

While a total lack of technical due diligence is not unheard of in such procurements, it was also a surprise to see the paper trail taken at face value in two other areas:

1. When reading a report, it's always nice to know a bit about the authors. We're not really familiar with the Gartner Group or the MIS managers who prepared the technical evaluations, so we prepared a little "getting to know you" presentation. While we've never been a Gartner client, we were surprised that Gartner did not disclose that it has had significant business relationships with NeuStar, VeriSign, and Register.Com.

2. "A demonstrated ability" was shown by many supplicants based on prior experience in the business. We looked around, but it appears that none of the reports that document actual performance results of the established players are online.

In short, it appears there's a bit of a traffic jam on the paper trail, but luckily there are several routes to our destination. After sitting in traffic for the last few months inhaling the exhaust from the money-guzzling vehicles driven by .commies and lawyers, we've decided to switch metaphors as a way of helping to save the environment. Outside the . beltway, fancy cars and loud meetings just don't have the same appeal. We look instead for our inspiration to the family farm. T.S. Eliot once asked, "what are the roots that clutch, what branches grow?" (Audio)

• Our roots are our current services operating on the Internet. You are invited to participate in our registry interoperability testbed, comment on technical notes, and examine our operational statistics.
• Our branches are the other people who use our code. You are invited to download our software for DNS and DHCP, and we are pleased to announce that training on our OpenReg open registration system will start early next year for our colleagues who run ccTLD and gTLD registries, as well as those that operate registries in corporations, universities, or other organizations that need to systematically delegate and allocate names.

In any case, the popular vote is in and all we can do is wait for the Electoral College to decide the future of the .org TLD. Thanks for your support.

Some ICANN Travesties for September 2002

ICANN launched a complaint against Verisign for having faulty woho is data. http://www.washingtonpost.com/wp-dyn/articles/A33395-2002Sep3.html
In addition to its complaint, ICANN today announced the establishment of a centralized online form (located at www.internic.net) that Internet users can use to report faulty Whois data. ICANN will also establish a tracking system to notify registrars of reported inaccuracies.

See

http://www.internic.net/cgi/rpt_whois/rpt

.cgi

Observer asked: Are there reasons to consider this a good thing? Can this/how can this be abused or used against registrants?

A second observer commented: how could it possibly be a good thing?

COOK Report: Remember this is the father of the internet making it safe for commerce by enabling a situation where if ICANN survives it can take someone it doesn't like OFF the internet for having faulty whois data.

Another observer: I'll go further than that. It's a huge move into governance. The Internic web site now says:

"ICANN does not resolve individual customer complaints. ICANN is a technical-coordination body. Its primary objective is to coordinate the Internet's system of assigned names and numbers to promote stable operation."

ICANN is following more or less the same policy as the FTC. The FTC are completely useless until the problems are serious enough for government intervention (i.e. very large scale). By then it's always too late. But at least they show up and eventually do something.

This situation now allows ICANN to bypass the FTC's consumer complaint process so the FTC (aka US Government) will never ever again know how bad the situation really is. There is also no verification or oversight of the data collected by ICANN. Obviously, this is a major step away from lawful accountability.

Later on BWG we read:

It has been a year since September 11, 2001.

What has ICANN done to protect the security and guarantee the recoverability of the Internet's Domain Name System?

- Staged a public tableau with many big names wringing their hands and saying the obvious.

- Formed a committee of big names. This committee seems to never have met.

- Ignored specific, workable, and easy solutions because they didn't come from "the right source".

DNS under ICANN is no more secure or recoverable today than it was at dawn, September 11, 2001.

From ICANN Watch

RIRs REALLY Mad At ICANN

Posted by michael on Sunday, September 15 @ 14:51:43 MDT
Contributed by michael

The three incumbent RIRs (the regional bodies that allocate IP numbers) have written an angry letter to ICANN.

So we have the end-users angry and powerless, the ccTLDs furious and powerful, and the RIRs furious, and sleeping giants. My guess is ICANN continues its pattern of trying to subjugate the ccTLDs and caving into RIRs at the last minute. But I'm not sure about that last bit.

Full text:

To: Vint Cerf, Alejandro Pissanty, Stuart Lynn

From: APNIC, ARIN, RIPE NCC

Subject: The ICANN Reform Process

This message is being sent to the Chair of ICANN Board, the Chair of the ICANN ERC, and the ICANN President and CEO. It is also being sent to the ERC comment list.

This is to acknowledge the teleconference that occurred on 10 September 2002. Participants were ICANN (Stuart Lynn, Louis Touton) and APNIC (Geoff Huston, Paul Wilson). While the APNIC participants were not speaking on behalf of ARIN and RIPE NCC, they were conveying the intent of the RIRs as stated in the two carefully prepared and coordinated papers commenting on ICANN reform. The APNIC participants of the call did not hear any indication that ICANN was reading these papers nor did they get any specific response to any of the points raised in those position papers. Therefore the RIRs make the following statement:

1. The RIRs are dismayed with the second interim report of the ICANN E&R Committee in terms of the lack of consideration and response to the submissions of the RIRs to ICANN.

2. In the light of this, the RIRs are unwilling to accept the proposed changes to the composition of the ASO, the proposed changes to the support mechanisms of the ASO and proposed changes to the role of the ASO within ICANN.

3. The RIRs will be providing a substantive blueprint for reform within the next several days.

Sincerely yours,

Paul Wilson
Director General
APNIC

Raymond A. Plzak
President & CEO
ARIN

Axel Pawlik
Managing Director
RIPE NCC

DoC Renews MoU

On September 23 John Paczkowski of "Good Morning Silicon Valley" writes:

"Commerce Dept. renews ICANN contract to mismanage Internet"
http://www.siliconvalley.com/mld/siliconvalley/business/columnists/gmsv/4134770.htm

A Short Case Study in ICANN Pathology

Editor's Note: Danny Younger may be a very nice guy. he has apparently tried to play ball with the likes of the ICANN Party Secretariate for a long time and Stuart Lynn graciously explains to Danny why so many folk like he get to see only the heel of ICANNs boot. Danny in turn plays the game by excoriating his ICANN "constutuency" underlings for not reforming their efforts. Stuart in turn may be very grateful to Danny for being able to say that unlike the ingrate law professor Froomkin there are folk like Danny who march hand in hand with the ICANN leadership. We note below that Milton has a different interpretation of Danny Younger's actions. Milton may well be correct. In any case this episode is one more example of how ICANN divides to conquer. We hope that the reservoir of people who still think they can affect the behavior of ICANN will soon dry up.

On September 24 Danny had the following complaint.

On 2 August, Jamie sent a note to this list requesting that someone be appointed in his place to serve on the Transfers Task Force. Seven weeks later Jamie again finds it necessary to reiterate his request to be allowed to resign from the TF. This should be a clear signal to all of you that a major problem exists within this constituency and that your AdCom is clearly not doing the job that they were elected to do.

I would think that most of you would be embarrassed by the fact that your own constituency website hasn't been updated since before Accra, that one can't even find either a list of your current Adcom members posted there or a list of your current paid-up organizational members, and that no recent Adcom teleconference minutes have been posted.

As an organization whose members are known for their penchant for criticizing ICANN reform, you seem to have done very little to clean up your own act and have apparently managed to allow your own constituency to fall into a state of total disarray and neglect. At the upcoming Names Council teleconference you will be asked regarding your plans to come into compliance with your financial obligations relative to the DNSO. What progress has been made on this issue? As usual, these matters haven't even been discussed with the membership, and it's doubtful whether you even have a viable constituency any longer.

The Blueprint recognized the need to retire constituencies when they are no longer viable. Ask yourselves if this constituency can continue to justify its existence -- you certainly can't even pay your own bills. You started out with almost two hundred member organizations -- how many now are current with their dues? How many members have you lost, and why have you lost so many members? More importantly, what are you doing about it?

How many remaining members actually participate in discussions of DNSO-related topics? What was the position of this constituency on the WHOIS report? Has the constituency even discussed the current Transfers TF report? Where was the position paper of this constituency on the ICANN reform?

Every other constituency sent an official constituency communique to the ERC's Forum -- yours did not. Why didn't your leadership articulate such a constituency position?

Reforming ICANN also means reforming each of its subordinate entities. You might want to start thinking about how you intend to reform your own constituency.

Milton Mueller, author of the scholarly monograph *Managing the Root*, a study that was too dispassionate for our tastes, finally had enough.

Milton Nails ICANN's Fraud

Danny:

"The answers to your questions are fairly simple. ICANN management has made it clear that it does not welcome public representation, via its abolition of the membership. The DNSO is a rigged game in which noncommercial interests are permanently marginalized. And the DNSO itself is marginalized - it can be and often is ignored by the Board. And the Board itself has no effective control of management. And the US Govt has just said it doesn't mind.

Scolding the members of this group for not continuing to invest substantial resources of time and money into this loser's game is, frankly, kinda dumb. Getting self-righteous about it shows that you have totally lost perspective.

Danny, you've played a constructive role in the past and I like you. But it's time to face reality. Volunteers like you who continue to pour resources into the current structure are suckers."

Editor's Note: At this point two paragraphs from Milton's original are struck at his request. Milton pointed out that further communication with Danny caused him to disagree with our interpretation. As he put it "Danny's problem is not that he is "playing ball" with Stuart Lynn or "marching hand in hand" with ICANN management, but that he really wants to affect how ICANN works, and he is frustrated that the voices of dissent (e.g., NCDNHC) have stopped doing anything.

Milton concludes "At any rate, we don't have to justify our existence to you or to the ICANN ERC. We are way, way, way beyond the point where threats of being "retired" can have any effect.

Maybe you should do a lot less ICANN and a lot more thinking about what is at stake for you and whatever group you think you represent."

Our thanks to Milton for permission to publish.

RIAA Runs Amok, Sues Four Backbones Over Chinese Site and Then Sues Verizon

Old Time Net Architect Makes Mistake with Affidavit in Support of First RIAA Suit Highlights

Editor's Note: On August 15 we had the extremely unusual situation of Dave Farber coming out on behalf of the RIAA and against the position of the EFF. In an affidavit filed with the RIAA suit Dave took the position that it would not be unreasonable for the court to order the four backbones being sued to block access to a pirate file sharing site in mainland China. He was saying essentially that in defense of copy righted content a court could order changes in how the Internet is run. This unfortunate action was something new under the sun and greatly alarmed us. Fortunately the RIAAdecided to pick on Verizon and dropped its backbone suit on August 21. Verizon, we were pleased to see, came out squarely in defense of its subscriber privacy as it file a brief on behalf of Kazaa and against RIAA. We chronicle the course of these strange events below.

On August 16, 2002 we wrote: **Dave Farber** has filed an affidavit on behalf of the position taken by the RIAA in a lawsuit against the major internet backbones seeking to force them to bloc access to an infringing provider in China. For the first time we have a person who participated in the construction Internet saying to a court that it is reasonable for it to order changes in the way the net is operated to satisfy the request of an owner of content. Let us examine what happened.

On August 15 a person on a private list had written: This might interest some: the record labels go to war with the Internet infrastructure. Fred [von Lohmann] is EFF's Senior IP Attorney. [He writes: to a public list]

I can't count the number of times I've seen posts on Pho speculating that infringers will relocate overseas, and thus escape the clutches of the RIAA. Each time, I've tried to remind everyone that the RIAA foresaw that eventuality and had 512(j) of the DMCA enacted, which allows them to force U.S. ISPs to block the IP numbers of foreign infringers.

Well, it's happened. A collection of major recording labels have filed suit against U.S. Internet backbone providers (AT&T, UUNet, Cable & Wireless, Sprint) under 512(j), demanding that they block the IP addresses for www.listen4ever.com (based in China). Action filed in Southern District of NY, preliminary injunction requested. Complaint and preliminary injunction motion available at: http://homepage.mac.com/fvl

The next day on Nanog, Richard A Steenbergen wrote to the list: Ok here's a question, why are they suing AT&T, CW, and UU? I see Listen4ever behind 4134 (China Telecom), who I only see buying transit through InterNAP. Wouldn't it be simpler for them to sue InterNAP? I guess it would sure be nice precedent, if they could make some big tier 1 providers do their bidding to filter whoever they want whenever they want.

Sean Donelan then replied: The problem with BGP is you only see the "best" path more than one hop away. The network in question is reachable through transit providers other than InterNAP, such as Concert.

http://www.nytimes.com/2002/08/17/business/media/17MUSI.html

The New York Times says the companies named in the suit are AT&T Broadband (not AT&T's backbone?), Cable & Wireless, Sprint Corporation and UUNet technologies.

"David Farber, a University of Pennsylvania computer scientist and an early architect of the Internet, filed an affidavit in the case, saying it would be relatively easy for the Internet companies to block the Internet address of the Web site without disrupting other traffic.

"It's not a big hassle," Mr. Farber said. "There's no way to stop everybody, but a substantial number of people will not be able to get access."

Cook: What is going on here? Here is Dave Farber, Internet civil libertarian, saying to a court that it is OK to order the big backbones of the net to change their routing on behalf of a content provider. What has earlier looked like common carriage now potentially becomes muddied by asking a court to get involved with the routing structure of the net on behalf of content. ICANN will love this. Why we wondered would Dave Farber do this?

Here is his answer: Farber to his IP list on Aug 17: "Since I provided an affidavit to this filing, I thought a bit of explanation might be illuminating. I have long resisted and will continue to resist the attempts of organizations to block access to the net in an attempt to restrict the freedom of speech and the fair use of information on the Internet. I am strongly opposed to any attempt to control the fair use access to music, text, pubs etc. This does not seem to [be] the issue in this case."

"This case was interesting for several reasons to me. First the site is an egregious example of a site that exists only to hold copyrighted music and offers a unreachable contact address and false advertisements in an attempt to look proper. Second, the aim of the case is the backbone suppliers not the local ISPs."

"I have always felt that the law we have on the books should be followed until the law is changed by the congress or the courts unless there is a much much higher ethical imperative that holds."

"This case seemed to be to be a good case to test the issues raised by DMCA (a law I believe was fatally flawed in concept and should be repealed) part dealing with copyrighted material on off-shore sites." [**Cook**: but Dave if you wanted to test the issues involved with the DMCA why did not you file an affidavit on behalf of one of the SPONSORS of DMCA?]

"I will be happy to send any IPer the affidavit on request. The affidavit addressed a set of technical issues and did so in a very limited context."

Dave

Cook: Before acting we would have hoped he would have gotten both technical advice on routing issues and legal advice on free speech issues? I see nothing constructive in this and plenty that is destructive. Consider the can of worms even on the technical routing side:

Sean Doran wrote on NANOG: Hm, why stop with just backbone networks?

Why shouldn't edge networks, corporate networks, and household networks chip in to uphold civil judgments against infringers? Surely I should not object if the RIAA insists that I block access of my small collection of computers to hosts which exist solely to distribute infringing material? Surely [M]r Farber has already volunteered to adjust his own border filtering, since it is no great inconvenience to do so, the case is clear-cut, and this bad law is still on the books and therefore should continue to be followed.

Sean Donelan replied: The record labels don't want to give you that choice. If you read the complaint you'll notice the record companies never attempted to contact the immediate upstream ISP in China. Instead of following international treaties for the service of process, which would take "months," they are forum shopping for a "less burdensome" (to them, more burdensome to everyone else) forum.

It is much easier to get on a "blackhole list" than it is to get off of one. If you are a non-US ISP, you could find your address space null routed by major US backbones without notice to you. Even if you later get rid of the customer, how does the non-US ISPget off the US court imposed blacklist? Will China Telecom, or the Chinese government need to hire a US lawyer to petition the US court for permission to have address space assigned by APNIC to China routed? Will RIPE and APNIC issue additional address space to a non-US ISP because their previous address space became unroutable due to US court orders?

Will backbones be expected to only null route addresses within the court's area of jurisdiction? Worldcom, AT&T, Sprint and C&W operate world-wide routing domains. How far will the US court's order "leak?" Will Canada and Europe still be able to reach www.listen4ever.com in China over the portions of the companies backbones not located in US jurisdiction? Likewise when German and French courts order backbone providers with assets in those jurisdictions to block access to illegal websites, how far will those orders leak? AOL/CompuServe has experienced this in Germany already.

Dr. Farber's statements to the contrary, I don't think this is trivial to implement. We have experience with AGIS, Napster, RBLs, etc.

Cook: I sent an expression of amazement to Dave Farber who replied presumably with regard to his earlier statement that "The affidavit addressed a set of technical issues and did so in a very limited context": Gordon, I do completely understand the limits of what I said and did so with complete understanding of those limits.

Judge for yourselves: Here is the affidavit which Dave sent as he promised he would:

Defendants' Routing Services Allow Internet Users in the United States to Access the Listen4ever Site

1. I understand that Listen4ever.com ("Listen4ever") has built, maintains and controls an integrated computer system and service accessible through a website http://www.listen4ever.com <http://www.listen4ever.com/> (the "Listen4ever Site"). Further, I understand that the Listen4ever Website is hosted on servers located in the People's Republic of China. I also understand that by using an easily used, web-based system, the Listen4ever Site enables users to connect to its central servers and encourages and enables them to download music from a centralized location containing millions of such files, thereby making unlawful copies of any and as many recordings they choose.

2. Cable and Wireless USA, AT&T Broadband Corp, UUNet Technologies, Inc. and Sprint Corp. – Advance Network Services (collectively, "Defendants") are in the business of providing Internet backbone routing services. Defendants provide routing connections to and from a website's host server, through which communications from a user's computer travel to the host server. Via Border Gateway Protocol ("BGP"), Defendants' routers exchange information about other known routers, the addresses they can reach, and the best available route between the addresses. Defendants' routers recognize addresses by reading the Autonomous System Number ("ASN"), or grouping of Internet Protocol ("IP") blocks, associated with the batch of addresses. The routers then determine the best pathway between the computer requesting a connection with a particular address and that address's host computer. Defendants essentially coordinate and provide the most efficient connection between computers on the Internet. Defendants' routing services are the principal means through which users in the United States can reach Listen4ever's servers in China.

Defendants Have the Ability to Block Access to the Listen4ever Site

3. Defendants readily have the technological capability to significantly limit access to Plaintiffs' copyrighted works

via the Listen4ever Site. By disabling the connections that allow users' computers in the United States to communicate with Listen4ever's servers in China, Defendants can significantly diminish the continued illegal copying and distribution of Plaintiffs' sound recordings. Defendants can arrange this blocking to ensure that only direct communications with Listen4ever's servers are blocked, and that communications with all other websites whose traffic flows through Defendants' backbone routing network proceed unhindered.

4. Defendants can block access to the Listen4ever Site in order to prevent copying and distribution of Plaintiffs' protected material by programming their routers to: (1) direct all traffic addressed to Listen4ever's ASN to an alternate site notifying them of the block, or (2) direct all traffic addressed to Listen4ever's IP addresses or IP blocks to an alternative site notifying them of the block. As it may be possible for Listen4ever to change its IP address, any injunction should also provide that if Listen4ever does so, then the new IP address shall also be blocked after notification is given to Defendants.

5. The processes by which Defendants can block access to the Listen4ever Site, as described above, are simple and technically well understood by Defendants. Blocking access to the Listen4ever Site would require Defendants to perform simple administrative and technical tasks, which would not place an undue burden on them or their networks.

Cook: Now the next paragraph of the above quoted *Times* article said:

The companies named in the suit declined to comment. But a person who works closely with Internet providers said that they were concerned about how easy it was for a Web site to change Internet addresses. If copyright holders began asking them to block sites in large numbers, and to keep track of every new address, it could divert resources from running regular Web traffic.

Cook: This is giving a court the opportunity to tell an ISPor big backbone get involved in content. I wrote back to Dave quoting his remark about understanding limits: "Gordon, I do completely understand the limits of what I said and did so with complete understanding of those limits."

Cook: I wish that you would elaborate.... on that above sentence. It makes no sense to me.

I don't think you made a good move....but I also need to be sure that if I criticize what you have done I understand what you were doing and why..... So let me ask...you talk of limits....what limits? Is it really your belief that what you have said can be expected not to extend beyond the "limits" of this case?

How can that be? Law is built on precedent.... should the court find against the backbone you have the precedent established that it is OK for a content provider to ask a court to block an IP number of something it doesn't like..... Why OK?....well this guy Farber, one of the architects of the net, says its ok ..... Where does it stop Dave?

I am not aware that a person from the internet side of the block and the civil libertarian side of the bloc such as you has EVER publicly and legally sided with the intellectual property folk....in any way shape of form....sided with them by saying the way the net is run should be changed to accommodate them. Now you have done so. Why?

How did you hear about this in time to do an official affidavit in time for the filing of the case? Are you an expert witness for the RIAA? It is your right to be such but shouldn't you disclose that if this is the case?

Given the positions you have espoused in the past, what you have just done makes no sense. I think you owe the net a detailed well-reasoned explanation. In my opinion the explanation you sent to your IP list is subject to severe criticism. As Sean Doran and Sean Donelan have pointed out your affidavit to the court left much to be desired.

The content owners are determined to break the end-to-end model of the Internet and grab control for themselves. Whose side are you on? Dave. Theirs or the Internet's. And why?

I am not aware that a person from the internet side of the block and the civil libertarian side of the block has EVER publicly and legally sided with the intellectual property folk....in any way shape of form.... by saying the way the net is run should be changed to accommodate them. Now you have done so. Why?--

Farber: I doubt I am the only one. Civil liberty is a lot lot more than the right to access some one else's music for free. To confuse these issues is, in my mind, to seriously damage the cause of civil liberty especially in light of the course the nation is going down.

As I said in my initial response, I believe that the egregious example of that Chinese web site both exceeds any definition of fair and provides an test case for the use of DMCA (which I have stated often is a bad law). I want the net to be the mainstay for future commerce and communications. If it is to be that it needs either conform to current laws and or get those laws changed.

Dave

Cook: Dave, are you thinking that we could have a legal precedent to take a web site off the air for infringing on music copyright and that this legal precedent would not function as a major risk of opening the door to demands from lawyers that speech of any kind be blocked because someone who can afford to sue doesn't like what is being said?

Civil liberty is inexorably bound up with free speech and the ability to publish... to create content and distribute content. The greatest revolution of the Internet was that it gave us all the ability to become publishers and content creators. As Lessig has so eloquently pointed out this is what enrages the content providers and intellectual property holders. Their monopolies are broken. DMCAand an asso-

ciated string of legislation including threats to change the design of the technology itself to give them the final and pre-emptive authority over content has been their answer.

Why should any of us assume that a court will say that an ISP should block a music trading website but not block a website with political economic or social content that a deep pocket litigant doesn't like and therefore harrassingly sues for infringement under DMCAor Sony Bono? What about the hundreds of small political discussion sites, where individual posters frequently extract paragraphs or even whole articles from major newspapers so as to be able to share and discuss them with other posters. Should all these discussion sites be monitored by cyber-air marshalls, or simply shut down in advance to forestall the copying of copyrighted news?

In my opinion control of content is the prize and has been the prize since 1997 when the GIP and other forces that put together ICANN coalesced. Legal precedent that enables courts to tell the operators of the transport layers of the protocol stack how to operate their part of the network to protect the interests of owners of processes that run at the application layers has been the game plan since the late 90's. Lessig in book The Future of Ideas (2001 has documented this strategy very well. Dave Farber has chosen to ignore them. He talks of a good test case against a bad DMCA. OK if that is his intent, why has he put himself on the side of the content owners and against the ISPs who operate the network. He admits that even if the backbones are told to block the site and do so it won't stop people from getting there but it will make it more difficult. Why has he joined the side that is determined to control the operation and the *content* of the internet? I have great difficulty making sense of his action. he says he wants to send a message that you shall not steal music. OK. But what he proposes may open doors to serious political free speech abuses and wont stop the stealing of music. It will just make the latter more difficult.

Now here is the final issue. he says: "I want the net to be the mainstay for future commerce and communications. If it is to be that it needs either conform to current laws and or get those laws changed." This troubles me. The GIP was started to make the net safe for commerce and not for free speech. The IANA was reformulated under ICANN for the same purpose. From 1997 through 1998 and beyond, the ISOC and ICANN folk talked about the need for adult supervision for the Internet. ICANN for sure has been turned into the adult supervisor. Vint in his Internet is for Everyone RFC pointed out that the Internet must be safe for commerce. Vint in his emails to ICANN staff and Tom Kalil at the White House in June 1999 said the Internet was in danger if ICANN failed. The same mentality as far as I can tell. ICANN's actions to date have been aimed just about entirely at helping large trade mark owners put infringers out of business. You look at the direction that ICANN has always headed in and the interests pushing it and the only conclusion is that it wants the right ultimate inspection and pre-emptive control of content on websites and elsewhere.

I submit that you cannot make the Internet both safe for commerce and safe for free expression. Yochai Benkler described it beautifully in a MAY 2000 paper: choose one - not both: the great agora (Athenian debating ground) or the great shopping mall. I want the agora. The one [Agora] or the other [shopping mall] must rule as the dominant principle of the internet by which everything else is judged, which principle will have to accept restriction when there is conflict between commerce and speech. I do not think that it is possible to accommodate the intellectual property folk and hope that there will be room left for free speech and civil liberties on the net when the dust settles. The intellectual property legislation we have seen in the last 10 years confirms that the IP owners control the congress, and that the only thing stopping them from literally becoming Big Brothers to us all, is our open internet architecture. I am genuinely surprised and perplexed that he seems not to see it this way. What he is doing is firmly aligning himself with those who have decided the internet needs customs inspectors of all content baggage at the center. While I think this is also Vint's Cerf's point of view I also think that the internet can run quite well on its own and as Lessig has pointed out there is a huge amount to be lost by adopting this stance.

The stance of siding with the music industry and the intellectual property owners will help to keep bandwidth demand down and will therefore also prolong the time needed for recovery from the industries' depression. A depression that is inexorably sapping the economic strength from software and hardware makers on the PC and server and data base and web-services side of the industry. Roxane **Googin** sees where this is headed. Larry Lessig sees it too. Dave as far as I can tell does not see it and thinks that you can compromise with these people.

I disagree. In protecting their expansive view of their economic interests they have taken and will continue to take action to defend their content and maximize everything else to extract the last nickel possible from it If Dave Farber's position prevails, the Pandora's box that it opens will not stop with this one Chinese web site. The ability for small content providers (yes I am one such) to use this technology to compete with the great corporate dinosaurs will be snuffed out. Inexplicably, Dave has walked into court and like Vint in the ICANN litigation has wound up on the wrong side of the bench.

On August 20 a **Nanog reader** asked: "if small (tier 4 - 5) ISPs can be threatened by its uplink for non compliance with the AUP (for example transmitting spam all the time), and medium ISPs (tier 3 - 4) can also be threatened by its uplink for non compliance with the AUP, then why tier 1 - 2 ISPs can't be threatened by RIAAA to comply to their AUP ?

Merit's **Jeff Ogden** replied: One difference is that there are business relationships between all of the upstreams and their downstreams. The contracts usually require compliance with the AUPs. If someone doesn't like an AUP they don't have to do business with that ISPand can

at least try to get service somewhere else. ISPs don't have business relationships with the RIAA and don't have the option to go somewhere else if the RIAA imposes its will on backbone ISPs.

COOK Report: Fortunately on August 21 the RIAAdropped its suit with a "notice of voluntary dismissal." Dean Garfield wrote to the defendents: You should have already received the notice of voluntary dismissal we filed today in the action related to Listen4ever - - a copy is also attached below. We are pleased that this action was resolved without the need for further litigation. We are also pleased that we were able to engage each of your companies in a dialogue on combatting piracy. In that regard, we are very interested in your ideas for addressing pirate sites like Listen4ever that choose to locate outside of the United States. As always, we welcome your views and insights on this issue in whatever forum each of you think is most appropriate. Thanks.

It was **Robert Berger** who got the issue really right when he wrote to Farber who also on August 21 posted the note to his IP list:

Both Dave Farber and Declan McCullagh are two of the strongest defenders of personal cyber-rights, privacy, fair access and appropriately limited Intellectual Property laws. They have always worked hard to act against the forces who have been chipping away at our rights in cyberspace (and physical space for that mater).

We have recent examples where they have also exercised the liberal tendency (I mean that in a good way) to speak out for the cases where they thought that the other side has been wronged even if it supported "the enemy's" position.

This behavior should normally be applauded even if one doesn't agree with them. It shows mature people who are more interested in what they see as truth than in politics.

What concerns me is that we seem to no longer to live in such a civil world. The people such as the RIAA, the fundamentalists and radical right all seem to be willing to forgo such niceties and will probably use these well meaning statements to only strengthen their position and will continue to use the constant repetition of their lies and any other technique to win and make others lose.

The right has been able to "stay on message" and set back the progressive clock by a huge amount. Liberals are still fighting among themselves and letting the right take the field without a real fight....

This is not good (IMHO)

My conclusion is that we should be really conscious of where we put our energy in the public conversation and maybe not be so generous when we see a legal point in favor of the people who have brought such a bad law to the books or start arguments that end up being just friendly fire.

RIAA Sues Verizon – Verizon Gets Creative

On August 20 RIAA sued Verizon seeking to for it to turn over the identites of its customers using the Kazaa peer-to-peer file service. Amazingly Verizon turned around and gave the RIAA a well-deserved slap in the face by joining Kazaa.

http://www.usatoday.com/life/cyber/tech/2002/05/14/music-kazaa.htm On August 27 USA Today wrote: An unlikely alliance of swap-service Kazaa and telephone and Internet giant Verizon is floating a proposal to break the logjam of lawsuits: Computer manufacturers, blank CD makers, ISPs and software firms such as Kazaa will pool funds and pay artists directly.

"Historically, there's been a clash between the content community and new technology, back to the player piano," says Verizon vice president Sarah Deutsch. "We're proposing the idea of a copyright compulsory license for the Internet, so peer-to-peer distribution would be legitimate and the copyright community would get compensation. It's hard to get the genie back in the bottle."

Our next issue (January 2003) will be out sometime in December. It will begin the VoIP Enterprise leaves the public network discussion. We leave for Nepal October 15 and expect to return November 10

Interview, Discussion, and Article Highlights

Intro State of Peering - pp. 1-5

Summary, Full Article

p. 1 Farooq Hussain has written for this issue a remarkably candid summary of the evolution of the Tier 1's peering policy. They are, he says, the Internet Core Networks that announced anonymously on December 5, 2001 their decision to move their peering to Equinix Exchanges. He identifies them as UUNET, Sprint, Cable and Wireless, Genuity, Level 3, Qwest, and AT&T. He also finds their peering requirements to be arbitrary beyond reason. For example, interconnection at OC48 is one thing, but to be forced to do so at 15 locations around the United States is something else again.
it.

p. 3 The tectonic plates of network traffic and power are shifting with the economic uncertainty brought on by the industry crash and the increase of cable modem and DSL traffic. Given the extraordinarily low cost of bandwidth and the existing investments of US carriers in some of the fiber players, we can expect very soon to see a build out on the part of these carriers into peering at Asian and European exchanges. In this sense a lot of effort will be put by large players into moves to enable them to avoid paying transit fees to the currently seven largest global backbones (Tier 1). In doing so, the likely outcome is that these new comers will eventually either replace or join the Tier 1 oligopoly.

They are, in effect, climbing a peering "ladder" where as their bandwidth increases and they peer with each other and can get peering with larger players, they are likely to deeper with smaller players whom they feel they now no longer need and believe they can sell transit to. Thus although the plates of peering are shifting the fundamental premise is likely to remain one where players peer only if their aggregate traffic is approximately equal. Bill Woodcock has shown that a case can be made that it makes sense for a larger network to accept traffic from a smaller network that terminates on that network. However the larger players are still firmly of the opinion that size differences in network traffic are there to be exploited by the larger and presumably more powerful network. Not surprisingly this view is firmly rejected by the smaller players. Therefore as the new broadband based networks move to extend their peering infrastructure around the US and across oceans, they are likely to act increasingly like the Tier 1 oligopoly they want to replace and seek to sell transit to rather than peer with others who haven't grown as fast.

What is unknown is how good a job how many smaller players can do of extending peering with other small players through use of the approaches and methodology outlined by Woodcock in the long interview in this issue. At an abstract level Woodcock's views that peering is a good thing seem to be understood and accepted by the large players as well. The problem for them seems to be one of "good for whom?" along with the belief that, once you get to a certain size (and we might add business model), they don't scale.

Therefore as the new broadband based networks move to extend their peering infrastructure around the US and across oceans, they are likely to act increasingly like the Tier 1 oligopoly they want to replace and seek to sell transit to rather than peer with others who haven't grown as fast.

Peering & Transit Broken, by FH pp. 7 - 11

Full Article

p. 9 In discussions with their European counterparts, the large backbone US ISPs have generally maintained a position that the balance of in-bound [from Europe to the US] versus out-bound [from the US to Europe] does not justify a peering relationship. European networks generally agree that this gap in traffic is closing particularly as content value in Europe grows. But there still exists a traffic disparity. Whether the traffic imbalance justifies a denial of peering in the US is open to question especially as the US networks, naturally enough, find it beneficial to maintain peering relationships in Europe with networks to whom they deny peering in the US. In South America and in the Asia Pacific region these imbalances are even more pronounced.

pp. 9 - 10 All this said, the central question for peering policy, taken on a global scale, as opposed to policy for the domestic USA, still revolves around who has to pay to get their traffic to the United States network infrastructure and who does not? And related to this one should ask what is happening within the US environment that as a consequence of recent bankruptcy and consolidation in the telecom sector might significantly impact the development of IP networks both within and outside the US?

In my perception as things are, seven networks [WorldCom/UUNET, Sprint, ATT, Level 3, Qwest, Cable & Wireless, Genuity] currently operate an oligarchy dominating peering in a manner detrimental to competition. A significant number of this group are in difficulties as businesses and may not survive intact for very much longer. Behind them are companies such as AOL and SBC who though close to being full members of this group are not quite there - yet. The impact of a peering oligarchy is a significant impediment to competition both for the domestic US market and internationally whether or not the size and composition of the group is marginally impacted by bankruptcy, consolidation or new members.

The structural framework for peering with these networks has its origins in the requirements set out for US national level IPbackbone networks at the time of

the NSFNET transition which I would argue are no longer appropriate especially as they have been used in the past couple of years to create a monument to restrictive practices - effectively eliminating all but the I-Core group from peering with one another. Equally, these I-Core networks have worked around peering and transit with "paid peering" relationships for certain networks while never formally acknowledging that this relationship exists. [**Editor's Note**: Paid peering is the ability to send all traffic from the network purchasing the peering that is terminating in the network from which peering is purchased and the agreement says that no transit is to be provided by the paid peer to the network purchasing peering.]

p. 11 However, at present most operators appear to managing a mix of transit, paid peering and peering relationships. In the mid 90s the incentive for US networks other than the three original backbones UUNET, MCI, and Sprint to seek zero settlement peering agreements, was driven as much by investor requirements that considered doing so essential to their ability to bring out an IPO as by that the cost of transit was always falling.

But for larger networks the costs of transit are a very significant component impacting their business efficiency. Some networks such as AOL for example might be able to leverage their market power to gain peering relationships or to otherwise drive down the costs of transit [If you don't give me peering, I take my dial access business elsewhere] but this type of leverage is not available to the vast majority of ISPs.

Woodcock Interview pp. 11-27

Full Article

p. 12 All that any ISP has to sell is the sum of all its peering. This is as true for the Tier Ones as it is for everyone else. We are talking about the sum of the outbound bandwidth. So for the Tier 1s it is only peering, and for the smaller guys it is the sum of the peering and transit. That being the case, someone who has restrictive peering requirements is simply not going to grow as fast as one who aggressively peers. They will just have less to sell and the less they have to sell the less money they can bring in from customers. Now fast growth on the part of those who peer aggressively doesn't mean that you still can't get ahead of yourself by borrowing too much money.

COOK Report: OK. I always had the impression that the successful business model was supposed to be one of keeping as much traffic as local as you can by doing as much peering as you can. But that ultimately as you go upwards in the hierarchical tree, sooner or later you get to people like the Tier Ones who say that if you want to get to the rest of the internet your non peered traffic is x megabits per second for which we will charge you y dollars per megabit per second to deliver.

Woodcock: That is exactly what the function of a transit provider should be. Someone to charge you some price for traffic that you either cannot deliver or don't want to deliver yourself. Zocalo never had peering in the Far East because, although we had a large volume of traffic that went there, it was never quite large enough to justify pulling a DS3 across the Pacific. It would have been phenomenally expensive. It was much easier simply to buy transit at the PAIX and hand off to a transit provider that did have that kind of connectivity and let them worry about it.

This is the kind of decision about economic balance that every provider has to make.

pp. 12 -13 The minimal ISP will have two transit providers. It will have transit from two different upstreams and it will have peering. Let's try stripping this down and see why it doesn't work if you have anything less than this. If you just have peering you cannot sell transit to someone else because transit means access to the whole Internet. You cannot peer with the whole Internet. No one in fact can peer with the whole Internet. Someone is always in some sense your upstream. If you are just peering then, you don't have connectivity to the whole Internet and you have to buy transit from someone. So you have to have a transit provider. But if you have only a single transit provider, what happens when that transit provider goes down? As it most assuredly will. It will either be them or the tail circuit to them sooner or later. To fulfill your obligations to always provide transit, you need redundancy.

Now let's say you have two transit providers only and no peers. The reason why that isn't going to work is because the transit providers have a certain value that they are selling at a certain price. If you are reselling only that and adding your own costs, the cost of your pipes to your customers will be higher than theirs.

COOK Report: What if your two transit providers were UUNET and Sprint?

Woodcock: It probably doesn't change things. My point is that if you are buying transit from two providers and you are adding your costs and your profit margin and reselling the result, you price to customers is going to be higher than their price to their customers. And your customer might as well bypass you and go directly to them. You have no value add.

To be able to sell reliable transit, you have to have two providers. In order to be able to have a value added so that you can stay in business, you have to peer. Because delivering of traffic locally at a lower cost than transit will reduce the delivery cost per bit of your traffic from what your upstream transit providers are charging you.

pp. 14 -15 This is going to sound kind of weird but when you get right down to it the Tier ones don't matter. They are insignificant in the over all scheme I things. There are too few of them. They aren't playing in the real economic space. Their dollars are not real dollars because they are being subsidized by other business units. They are not profit centers within their companies. They are not doing anything to lead the market in new directions.

COOK Report: so the point is that there are so many Tier 2 possibilities for rout-

ing traffic that Tier one stranglehold are in danger of being broken? Perhaps this is one reason why when E-bone was shut off a few weeks ago its traffic was absorbed with barely a hiccough?

Woodcock: Right. I am not saying that if the Tier one backbones were turned off there would not be a huge effect on the Internet. There would be. But I am saying that in terms of guiding the development of the internet business model as a whole, they are not major players. They don't lead by example. Because their income comes from voice minutes, their Internet business plans don't even have to be self-sufficient.

COOK Report: They just do their own thing in a vacuum?

Woodcock: Yes. They are not really affecting what anyone else does. There is no exchange point in the world that would be adversely affected if the Tier ones disappeared tomorrow. They simply don't have a major affect on the peering infrastructure which is why they don't make a major difference to me.

What we are seeing in looking at exchange points is a differentiation between peering exchanges and transit exchanges. Peering exchanges are really cheap and quite large because there are two ways that they become worthwhile. Either they can be really cheap or have lots and lots of peering or preferably both. In a given region you want to go to one exchange and peer with as many people as possible. A second peering exchange in the same reason is actually not a good idea.

p. 16 For example just because an exchange is there does not mean that you too need be there. Even if an exchange is clearly within your services area you don't need to be there unless you can save money by being there.

COOK Report: Who does the modeling? It sounds top me like there may be a service business here for someone.

Woodcock: Yes quite possibly. If you look at the modeling Stephen Stuart did for Abovenet and MFN, it parallels the modeling that I did for Zocalo and that Avi did for Abovenet. I think that a fair number of other ISPs, ones that were profitable and careful with their money did exactly the same thing.

What you do is to turn on Netflow in all your routers and you log the bit counts that are flowing through them to every other AS in the known universe. And you rank order your traffic. You see where you traffic is going and you make a 'hit list'of the destinations. For example number one on such a list might be Sprint. Or more likely Cable and Wireless. Are either going to peer with you? Most likely not.

But there are two ways of looking at the list. The way that occurs to most people is to look at the adjacent peer. They say to whom do we connect right now that we are sending the most traffic to and how can we reduce that cost? The better way of looking at it takes more code, but a fair number of us do it this way. is to take the entire AS path between you and every destination and then you allocate points to each AS based on the amount of traffic that could potentially go through it. And then you must have a knob to help you determine how strongly you weight in favor of short paths. And you twist that knob back and forth a little bit and find that it changes your ranking. Suddenly you may find that Cable and Wireless, UUNET and Sprint are no longer among your top three. Pac Bell DSL, Chinanet and Road Runner might pop up at the top. What is significant is that there are ways to peer with folk like them because they are also interested in bypassing Sprint, Cable and Wireless and UUNET.

Your objective is to look at where your are sending traffic from the point of view of finding ways around those very large next hop destinations. It is not where you are sending traffic right now and who is already next to you. You want to reduce the amount of traffic that you are sending to places right now. The goal is how you can get around them and to reduce the amount of traffic going to them.

p. 20 Tudor: An ISP needs to be continuously concerned with lowering its per bit cost. The following methodology (credited to Bill **Woodcock**) quantifies an ISP's usage of its resources, namely peering & transit connections and suggests possible changes. The numbers produced together with an ISP's actual cost can then be used for an actual monetary calculation.

Quantification of resource use simply means the amount of bi-directional traffic - volume in bytes - an ISP exchanges with its peers and transit providers. We differentiate between two traffic types: transit and terminating. Transit traffic passes through an AS; terminating traffic ends in an AS.

For the purpose of explaining the method, let us assume an ISP with one router and several interfaces connected to several transit providers and several exchange points. From the router we collect: a) per prefix aggregated flow volume (using Netflow) and b) a full routing table (RIB) snapshot.

We perform two sets of calculations, one using the selected AS path as reflected in the RIB (reflected in Tables One and Two), the other synthetically constructed (reflected in Table Three in two parts below). The meaning of 'synthetic AS path' is defined later. First, the 'real' AS path.

For each prefix we have multiple alternatives, depending on how many transit providers we have. For example prefix 10.0.0.0/8 may be offered by provider 1 via AS path '1 23 100' and provider 2 via AS path '2 6 100'. The router however will make a choice between the two offered paths and use it. This choice is reflected in the RIB for each prefix. Thus, for example, if our flow data shows 1000 bytes for prefix 10.0.0.0/8 and the chosen AS path was from provider 2, we will attribute 1000 bytes of 'transit volume' to each of AS 2 & 6 and 1000 bytes of 'terminating volume' to AS 100. The result of this daily calculation yields the following output. (See **Table One** bottom of next page.)

Woodcock: He is using the Netflow data which gives you bit counts for source and destination AS prefixs. You know how much data is going from you to any destination AS prefix which basically is one of about 112,000 IP subnets to which there is a route advertised in the global routing table.

p. 23 Woodcock: In these final two tables we have synthesized a composite AS path. This is saying that we are not looking at the AS paths for selected routes but that we are looking at ALL of the potential ones – everything that we are being offered.

In the first two tables we were only looking at the subset of the routing table that we were actually using. In the first one we were receiving routes from everyone that has an asterisk. But we are not necessarily using every route that we hear. We are picking based upon a BGP selection algorithm, one route to every destination. The first table used our actual routes. But it is neither as thorough nor as complete as it could be, because we have additional information. If our BGP selection algorithm were more clever, it could hypothetically deliver traffic in different ways to different peers. There are a lot of other criteria it could be using. Or it could be apply the same criteria, but in a different order to get a different result. What we are doing here is saying that we were offered five different ways of getting to this destination.

Instead of just looking at the potential of peering with people who happen to be in the path that we actually used, as a way of shortening the path, what if we peered with someone who was in the middle of the path that we were not using before? Obviously it would not be a win for this particular destination. But if we peered with that entity, it might be that doing so would bring us closer to a whole lot of destinations for each of which we have some traffic. This is saying that while someone might not be the greatest peer in any particular case, it might be that peering with them would bring us a little closer o a lot of the Internet as opposed to a lot closer to some specific part of the Internet.

The real BGP decision-making algorithm that is applied here is myopic. It looks at one prefix at a time. It asks what is the shortest path (fewest AS outer hops) to this particular prefix. It ignores the amount of traffic. It ignores dollar costs. It ignores a lot of things. One of the things it ignores is how close and AS is to other destinations that it doesn't care about for delivery of a specific packet. BGPdelivers by whatever shortest path it has been given. It doesn't care that a packet may be deliverable via a route other than the one it has been given, it follows only the route it knows about, because this analysis is not looking at the specific path chosen by the BGP selection algorithm. It is looking at all the paths. And it is creating a synthetic path that contains in an unordered list everyone of the ASs that could be between us and the destination. Not the ASs that were actually between us and the destination in the path that was historically chosen by the router at the time the packet was delivered.

Contributor's Discussion pp. 28 - 79

Full Article

p. 30 My name is Roxane [Googin]and I am your worst nightmare. I care about asset allocation and how technology changes impact the overall economic scene. For about 2 years now I have been of the belief that IPtechnology represents a paradox. It is both our pathway to the future and economic kryptonite. It is so good, it is unfundable. You can read about my thinking in "The Paradox of the Perfect Network." See the Isenberg/Weinberger write-up at <http://www.netparadox.com/>

Bottom line, IP is such a perfect commodity, it guarantees its suppliers a loss on operations. For its job, of communications, this makes it perfect, because it neither imposes any preparatory restrictions on communications nor does it run out of capacity. Thus, the maximum creativity at the edges means zero value-add for the middle. Think of it as preserving entropy or something. This also means that no one makes money in the middle. This means our future well being is dependent upon something that no one benefits from doing.

It also means that all telcos on the planet go broke, taking their investors with them. This is where it gets messy. How do we get out of our legacy investments? Who pays for the next build-out? (Hint: it has to be social as the markets are not about to touch this one) This is a non-trivial question. The entire market decline is based on this problem. The crooks, the bankruptcies, the lack of revenues all stem from this one truth. The legacy telcos have about $1T in debt globally. It is worthless. Those bond holders will never get paid back. When do we admit this and quit waiting for the "bottom" or the "second half rebound"? There is not one. We just go down until we really crash. This is just a warm-up!

Then, how do we convince our government that the telcos are in fact the bad guys, and the crooks who stole money actually run the right stuff? For IP to win, someone other than the legacy Telco guys must manage the network. Its buildout cost must be sold to an increasingly impoverished public. The only guys who have not embezzled money are the ones who cannot be allowed to survive, because they will do anything they can to prove IP is useless and SONET is all there is. If SONET lives, our economy dies, as we cannot move to the next productivity paradigm of the real-time organization using Web services on anything other that ubiquitous, Gb Ethernet. If we have that, the legacy telcos are broke. It is that simple.

My bottom line question is: how much will a "good enough" buildout cost? How can we sell this to the Government as simple enough to manage they are not creating another postal service or Amtrak?

p. 40 ***COOK Report***: LINX is a switching fabric connecting exchanges predominantly used for transit?

Mike Hughes: Modulo Keith's comment regarding the "exchanges" - i.e. most being neutral co-lo facilities, selling

"housing real estate", with fairly minimal value-add services (most commonly remote hands/facility management) - LINX connects 10 different co-location buildings. The buildings themselves are used for all sorts of tele-housing, ISPs, web-hosting, some disaster recovery, enterprise computing, telco nodes/central offices.

COOK Report: If I am buying transit at Telehouse, I can buy a membership in the LINX switching fabric which I would use just for peering?

Hughes: You don't have to be buying transit at Telehouse. You could just be in one of the buildings where our switches are, one of which happens to be Telehouse. Though, a lot of people who do have a presence in Telehouse tend to buy transit there too, because it makes sense to do it - no tail circuit, good choice of readily available carriers.

COOK Report: What is not quite clear to me is whether joining LINX gets one peering with everyone else?

Hughes: Nope. No MLPA here! You negotiate the peerings yourself, bi-laterally. Fortunately, most LINX members have a fairly open peering policy! Many don't even require a peering agreement/contract, a "handshake" will do.

pp. 42-43 *COOK Report*: But is the reality smashingly complex? Or is Bill really saying in the midst of all this smashing complexity here is one way you can gain some control and comparative simplicity?

Hughes: Going peering gives you a lot more control of how your traffic reaches your network, or goes toward it's destination, if that's what you mean? Exchanges and neutral carrier hotels take this one step further by getting all this under the same roof!

Mitchell: I think like many realities, it only looks complex until you understand the underlying abstractions that are driving it all. Better understanding and communication of these abstractions is what many of the people on this list have (are !-) been grappling with.

Woodcock: I guess I tend to view the routing, the data-path, and the money as three separate but interrelated layers. The parameters do indeed seem fairly simple to me: the goal is always to have at least one route which is associated with a usable forwarding path for any destination IP address, while minimizing cost. Since failure of any system has a probability of 1, most of the work goes into trading off just how many routes you feel like carrying, against the cost of doing so. This all assumes that one is optimizing along logical business principals, not just trying to keep problems minimally visible while optimizing share price.

So given that, I basically get back to the minimal ISP reference model that I proposed: An ISPis someone who buys transit from two sources, peers as much as possible, and sells transit to customers.

You obviously negotiated the lowest transit price you can, from within the set of prices available from providers who you consider capable of making a commoditized "full routes/usable forwarding path" service actually work, and then you figure out how to minimize the amount of it that you have to buy, while maximizing the number of modulatable bits you have to sell to customers. Which means constantly reevaluating the peering sessions and places you peer in light of your real-world traffic mix.

Which is really the crux of the work in our industry, I think: the algorithm for determining the economic threshold of viability for participation in an exchange. Everything else falls out of that decision as a consequence.

pp. 44 -45 Woodcock: Alex's posting includes pseudo-code for a description of the model we put together for this a couple of years ago. And my assumption is that it's functionally similar to the one that I know Stephen was using for AS6461.

Basically, take your Netflow exports and the set of all ASes which appear in any AS-path associated with each destination prefix. For each bit to a destination prefix, apply to a bucket associated with each AS one point times a variable divided by the distance at which that AS appears from the origin of the path. Tweak the value of the variable high to indicate that you favor aggregation (small number of transit providers/peers, longer distance to each destination) or path length (large number of transit providers/peers, short distance to each destination). Refine by discarding anything in a path that's to the left of a known tier one.

That gives you a comparison of the relative merit of any AS as a potential peer. Redo the calculation with just selected paths, no weighting, and factoring in cost of delivery, and you get actual per-bit delivery costs for all your bits. Modify your routing table with information you glean from a looking glass to simulate your new selected paths after a hypothetical new peering session has been added and re-run, then figure back up to a total cost, to see whether you should actually be doing any specific peering session. That seems pretty concrete to me. Stephen, is that about the same algorithm you use, or do you do something different?

Stuart: Well, the goal was somewhat different than what a typical ISPwould have in mind; often what we were trying to accomplish was to reach a certain traffic ratio with a specific peer, or to reduce traffic overall to/from a specific peer. The structure of the algorithm would seem to be basically the same, though; just with more initial conditions and some slightly different tests.

The focus on economics, though, misses a point of our peering philosophy that is probably not in vogue in today's economic times: peering increases the quality of the network. When we peered with a network, it introduced a direct path between that network and its customers and our network and our customers, such that performance problems could be addressed in a one-on-one fashion as dictated by the peering contract - networks that we reached through peers were subject to performance problems in distant peering/transit connections that affected our customers but where we had no contrac-

tual leverage to cause to be fixed. Qualitatively, I can say that difference was important in a couple cases. In economic terms it probably increased the cost of peering (since peering wasn't approached from a strict economic perspective), but it also retained customers.

p. 49 ***COOK Report***: Do any of you have any opinions on whether Bill **Woodcock**'s views on peering and transit and their purported effect on ISP viability have anything to say about what economy of scale in the Internet is all about? Surely Fast Net has a very very different economy of scale and hence viability than SBC?

Freedman: Yep. FastNet can become profitable and SBC can't as easily. Or maybe they both can become profitable.

There are two factors working:

1) Large companies tend to do things less efficiently. Along the curve of wasted overhead/revenue the less than 10million per year companies and the greater than the billions of dollars companies typically do the best - but at current DSL prices, with infrastructure costs and assuming capital is not free, I don't see how SBC can make money (i.e. price can also prevent companies with revenues in the multiple billions from making money).

2) In the peering game, certainly there is more efficiency the more traffic you have, if you can keep the provisioning etc groups sane internally in terms of overhead. Of course, at a certain size the infrastructure is a killer, as 10gb/sec ports are much more expensive than 4 x 2.5gb/sec ports. But in general, 100mbit/sec of transit is cheaper to sink via peering than .5gb/sec is and up the line. It mostly has to do with the efficiency of local loop and interconnect costs.

If you put all of your eggs in one basket and everyone moved to the Equinixes, our Internet robustness would suck (re: resistance to attack), but the extra benefit to scale would be less, though still there.

p.53 Klein With some financially troubled organizations actually turning whole networks (or just segments) down, could this be constraining transmission resources in some areas? Putting whole fiber rings, PoPs, etc., into limbo until the administrators dispose of the assets? Has anyone done any research into how deep this has to cut until it becomes visible in the market price (not cost, note) of IP transit or clear circuits?

I guess this needs to be looked at from two angles as well, one is hub-to-hub capacity where there is plenty of provider choice. The other would be in the more fiber-remote areas where there is currently a small oligopoly of providers. The withdrawal of a small number of providers from those markets leaves only the RBOC/incumbent PTT and maybe one or two competitive carrier(s). Is that enough to force prices up in those markets (as long as there is maintained demand)? Just a thought.

Woodcock: This actually brings up a different concern of mine that's been bugging me more lately. The fiber that we're all using right now was, for the most part, financed by bankers who were looking at amortizing it with some amount of high-dollar-value voice-minute traffic occupying the extreme bottom end of the bandwidth. All the excess capacity then got sold off at whatever price could be gotten, for Internet use. That excess capacity, in turn, is getting sold off to VoIP providers, who are undercutting the $1 per minute traffic at $0.03 per minute. So what happens when we use up the current capacity, and have to explain the amortization basis for the next round of installs to bankers?

I'm not explaining this very well, but it looks like it might be a relatively large problem five years or so from now.

Spenceley: This mirrors similar concerns I have for the AP region, with the wonderful increase in capacity we have seen in the last the years and the availability of long-term IRU's. The corresponding increase in capacity that has essentially been sitting idle, coupled with the pressures to increase revenue from backers/boards/shareholder et.al. making executives and sales types come to the wonderful conclusion, they have such spare capacity its better to sell it below cost than have it sit idle.

pp. 56 -57 My strong impression is that while the costs of transit have plummeted, the prices paid by the end users (at the T1 and sub-T1 level, where the bulk of the revenues are) have declined much less, and of course there is a far greater density of them.

Woodcock: That's certainly my impression as well. Bulk transit costs seem to have come down from perhaps $800 megabit to $100 megabit over the last four years, while retail has come down from about $1200 to about $550 over the same period. Does that square with other folk's general impression of the numbers?

I wouldn't argue that that indicates greater profitability, though, by any means... Just that people are dumping excess capacity to big customers who can chew it up quickly, at whatever they can get for it.

Costs Move from the Center to the Edges

Odlyzko: I was certainly not implying greater profitability, although that may also be true. This is probably the result of the natural evolution of the industry, with costs moving towards the edges. We can see something very similar in the computer industry. The power of the leading-edge microprocessors has been increasing for several decades at about 60% a year, as described by Moore's Law. Their prices have stayed stable, at a few hundred dollars each. On the other hand, the prices and computing power of 5 MHz 8-bit embedded microprocessors have not changed all that much.

Freedman: I think retail has come to $180/mb to $350/mb or so, but not $550/mb. UUNET will quote you $400 per month for a t1 (full) or $225/mb for collocation, I think.

Hussain: I mostly hear much lower numbers for paid peering between $50-100 per meg. It varies obviously depending on who is selling but numbers as low as $25 per meg for paid peering are ru-

mored though between $50 and $100 seems more the norm. These deals may be cut with reciprocal arrangements elsewhere for purchase of capacity. It would be really valuable to have some transparency in these numbers but I fear that is unlikely to happen.

Nowlin: Paid peering often takes the form of one party purchasing all the loops/cross-connects to the other party. In extreme cases (1998 time frame) there were cards for routers and even fiber discounts in exchange for peering. Still paid, but more difficult to pin a dollar per megabit charge on. Loop or cross-connect MRC covered by one party of the two peers is still the easiest way to deal with the downward spiral of per megabit charges when offering up a middle ground to peer/no peer negotiations.

Odlyzko: Along the same lines, transit revenues from ISPs that can get the $150/Mbps per month prices from Tier 1s simply do not come to all that much. Most of the money the Tier 1s make comes from end users. That should be kept in mind when evaluating the tactics those guys use in setting their peering policies.

There are some other basic assumptions in this discussion that I have doubts about. For example, in his first interview with Gordon, Bill said "It probably doesn't change things. My point is that if you are buying transit from two providers and you are adding your costs and your profit margin and reselling the result, you price to customers is going to be higher than their price to their customers. And your customer might as well bypass you and go directly to them. You have no value add."

Is that really so?

pp. 64-65 Doncaster: The claims of a $100/mb floors would seem to be 2x higher than reality.

Here's a clip of a post made by kirk@wolf.net to isp-bandwidth a couple months ago. He's apparently an independent sales agent for a number of carriers. Note OC12 pricing from Sprint & WorldCom work out to under $50/mb. Adding a 0-mile loop in a place like 60 Hudson or 111 8th shouldn't increase that by much more. I've personally been quoted $100/mb for 50mb (95th percentile) burstable IP transit delivered over FE, from QWest in 60 Hudson. Additionally, I have evidence (under NDA unfortunately) that indicates Level3 is selling burstable IP transit to customers buying multiple Gig-E ports for $<40/mb.

OC3 Pricing:

Sprint - $14,600 + Loop (Limited Areas)
NTT / Verio - $8990 + Loop (Cisco Router Provided Free)
Global Crossing - $12,000 + Loop
Genuity - $20,000 + Loop
MCI / WorldCom - $10,000 + Loop

OC12 Pricing:

Sprint - $30,000 + Loop (Very Limited Areas - not for pricing ... for= availability)
NTT / Verio - $24,880 + Loop (Cisco Router Provided Free)
Global Crossing - $32,000 + Loop
Genuity - $42,000 + Loop
MCI / WorldCom - $30,000 + Loop

COOK Report: Any comments on this? Are the carriers so desperate that they have independent sales agents?

Freedman: Yes. Carriers are all desperate.

COOK Report: Ralph, In your above example are you guaranteed 50 megabits burstable to 100 if the bandwidth is there? You are paying 5,000 a month for 50 meg Ottawa New York? Is that really a "deal"? Here the floor appears to be 100 bucks. No?

Doncaster: The QWest quote is $100/mb, yes. The point is that it's for ONLY a 50mbit commit. And the commit means I would commit to paying for 50mbps (95th % measured) even if I didn't use the full 50m. At any time I would be free to burst up to 100M, but if my 95th % was above 50m for the month I'd pay an extra $100/m for the difference. I didn't take the QWest quote though, I went with a better deal that gets me as good (or I think better) quality transit for 1/2 that price.

As for the cost of my OC3 to Canada, it's less than $2,000 per month and it's to Toronto 151 Front St. However that's not paid to the IPtransit provider, and I don't HAVE to back-haul the IP to Canada, I could easily sell some of it to potential customers in 60 Hudson.

Freedman: Well, one thing to note - $48/mb at full OC12 rates is really like $100/mb 95th - probably more like $120/mb - on a burstable pipe.

Doncaster: 95th percent compared to the peak on big pipes like OC12 is typically a 70-80% ratio. So 48x1.33 would be a better equivalent, assuming you know how to twiddle with your network to keep peak near 100%. In other words, get the Sprint OC12 and burstable Gig-E from another provider. When the Sprint OC12 nears congestion levels, shift some traffic over to the burstable Gig-E. Now if you aren't good at managing traffic I'd agree that you could expect the 95th% on an OC12 to fall around 300-400Mbits/sec.

Freedman: I disagree, but it could be based on different traffic profiles. Many of the oc12-size pipes I've seen all roughly bill 50-60%ile off of peak, especially if you have event-driven customers, though there are fewer of those nowadays.

And average is almost 50% better than that. But Your Mileage May Vary and that's just what has worked for me. The main point is that $50/mb flat-rate can't be compared to $50/mb 95th percentile without a conversion factor..

p. 67 Nowlin: I agree with Joe [**Klein**]. Cable folks have taken a serious chunk of traffic away from the 'tier 1'club as dial-up users are canceling the second phone line in the house to save a few bucks. Measurement tools in use at SBC confirmed my gut feeling about this trend. We actively pursue cable networks as peers (hi Joe!) because they are typically

behind one of the tier 1 club members and are growing much faster than the traditional transit suppliers. Telecommuters often use cable access if DSL is not available (I'm on RCN all day long) so it isn't just the nights/weekends that add to their traffic now.

Another reason why cable peers are good to pursue is that some tier 1 club members still enforce traffic ratios. Take the content out of their path. Many of them are in tough financial positions and are not adding capacity where they should. Some of the measurement tools available show they are loosing steam as customers do not renew long term agreements with such unstable vendors.

Most on this list are familiar with why a 'tier 1'wouldn't peer with SBC (multiple ASNs) even when we are able to meet them at more than a dozen common foot prints so forgive this backgrounder here for the rest. When I started at SBC nearly a year ago much of what exists behind AS7132 today was in a variety of ASNs for regulatory reasons and only seen via transit. SBC uses Sprint to carry IP traffic where LATA boundaries are crossed in states where local market-opening requirements of the Telecom Act's Section.271 have not been met.

p. 68 Nowlin: For 6+ months I've been doing weekly top 10 in/outbound, top 20, top 30, top 40 & top 50 checks using Adlex traffic tracker. The Tier 1 club should be separated into growing versus shrinking camps. Only four networks from that club remain in our top 10 checks. Sprint, AT&T, UUNET & Level 3. The rest who claim it fall anywhere in the range of 20-60 now depending on inbound or outbound flows." Some other sources are adamant that while this data may be true for the source's network, it is not universally reflective of what is from other networks that track the traffic of these seven.

Woodcock: Ren, how long ago did C&W fall out of the top ten, and how far down are you seeing them now?

Nowlin: Around the same time Genuity did, so probably not long after much of the Exodus network was dismantled and customers started to run in other directions this summer. They are in the top 20 in some of our regions, top 30 in others. Some of AOL's RoadRunner holdings are higher in rank than C&W... Of course AOL's AS1668 is always in the top 10.

pp. 77-78 Stuart: I'm of the belief that a significant amount of the content that would drive broadband sales, backbone utilization, etc., is locked up in RIAA/MPAA stupidity. Delivering that content could be the next "killer app," but the RIAA/MPAA goons are so focused on getting every bit of revenue for themselves that the orifice remains tightly closed.

Diaz: I agree that if we could release more of the content out there, it would drive more backbone traffic. At the same time, some 'killer apps' would be created. If we can consider that video bandwidth or peer to peer is a killer bandwidth app, then we could see backbone growth once again skyrocket.

Let's face it, I have a gigE built into my MACs now, what difference does that make when I'm stuck with a 786k DSL line, or a paltry 10meg in the office environment.

I think the reason why more technology like remote apps or remote desktops hasn't taken off is that it doesn't work wide area on these sized pipes. If we could increase the pipe size, then these services would take off — possibly driving additional growth.

Odlyzko: Yes, making things like music would increase demand for broadband. In the case of South Korea (whcih now has by far the greatest residential broadband penetration), interactive games and social factors, together with low prices, seemed to be crucial. (You might like to look at Izumi Aizu's paper "A Comparative Study of Broadband in Asia: Deployment and Policy," <http://www.anr.org/web/html/output/2002/broadbandasia522.htm>.)

On the other hand, you could not do everything at once. At today's prices for transit (say $150 per Mbps per month) shipping a DVD of about 4 GB will cost you about $1.50 in transit costs alone (and only if you run your connection flat out), so in practice you would have to charge something like $10 or $20 per DVD. (Peter Wayner's column in today's New York Times, "The packaging of video on demand," <http://www.nytimes.com/2002/09/23/technology/23NECO.html>, talks about this.

"Do ATM-based Internet Exchange Points make sense anymore?", pp. 80 - 83 Full Article

p. 81 Bill Nortin; In this paper we apply the peering financial models to this question, using current market prices to compare the price of transit against the costs of peering at ATM-based NAPs and Ethernet-based Internet Exchange Points. We build upon the previous research on Peering by introducing the notion of an Effective Peering Range (EPR) to describe the "useful life" of an Internet Exchange. We also highlight a potentially costly EPR Gap, an interim range between Peering Capacity points where peering is more expensive than transit.

The financial models presented that produced the graphs are included in the Appendix so that ISPs can apply these cost models to their specific situation.

Cost Reduction from New Semi Conductor Technology, pp. 84 - 88, Full Article

p. 84 Googin: Some may look at the cost of LEC infrastructure as a barrier standing in the way of broadband There are significant changes afoot above and beyond fiber that are going to knock the foundations out from under the price of LEC infrastructure.

One of the big cost reductions that we are going to get will come from semiconductor technology. In order to get to 10 gigabits you generally use silicon germa-

nium or gallium arsinide or some other exotic material. The ASPs are high. But the semiconductor industry is now going from aluminum to copper. Copper will make things faster. It is also moving from .18 microns to .13 microns. At .13 microns you have changes in the dielectric layers between the metal runs.

You have an issue with capacitance. It is like a battery. When you have two plates and some material in between - even air- you have energy storage. It functions like a battery as it captures the electrons and stops them from moving. The power of a capacitor goes up as a function of the inverse of the square of the distance between the plates. As lines get closer, it goes up fast.

p. 84 These changes will have a profound impact on the cost and performance of transport equipment and especially metro transport. AMCC had been using gallium arsenide and have switched to become 80 per cent CMOS. What they both have said very clearly is that if the speed of telecom transitions had kept up and we were moving on to 40 gigabits, you would need the gallium arsenide. But since the telecom industry has stopped in its tracks for a technology generation, that CMOS has now caught up. Now you can do 10 gigabit physical layer chips, framer chips and network processors in CMOS rather than in the more exotic gallium arsenide.

p. 85 For a 10 gigabit switch you have your physical layer chips. The "phys" from AMCC that have to be the drivers. Then you have your framers which are different chips. These put your packets into frames and do your addressing. Then they have their network processors that they think of in layers of intelligence. Not only by the beginning of next year can they do the individual sections in CMOS as opposed to fancy and expensive gallium arsenide, but then they will also start to integrate all of these into one chip. You might want to guess how many gates it would take to put an entire phy, and framer and network processor onto a single chip that would cost about $250.00. That is the sort of price you should be thinking about for these things. It goes from millions of dollars to $250. A switch on a chip. That is the power of this transition. It is driven by PC economics. The finished product should be about as expensive as a PC.

pp. 86-87 ***COOK Report***: So these new and cheaper switches will be set up to run SIP and Voice over IP proxies?

Googin: Right. A hard part about this transition is that it is not just a rebuilding of the old stuff to run better, faster, cheaper. It is building something new and different. No one but the monopolists are going to miss SS7. The developers hate it. I still don't think the software quite works which is a really interesting thing to start focusing on.

Now LEC CAPEX is down 40% compared to a year ago. In 2002 first quarter Verizon sold five billion in new bonds. Every deal was over subscribed. The interest was treasury rate plus 1/2% which indicates the perception of low risk.

COOK Report: As long as they can do this they can postpone the end by doing a kind of refinancing.

When Refinancing No Longer Works

Googin: Indeed. This is what they have been furiously doing for the past year. But all of a sudden the music will stop and when it does it stops really fast. Something happens and people get scared. That's how the debt markets work. They are really nice until they won't talk to you. Look at what happened to Qwest. They had a problem with their commercial paper and suddenly the company was weeks away from bankruptcy. Now they can't raise any new money. These debt guys aren't real swift. They are nice until they lose money, then they are GONE.

The first one that Verizon did was a billion dollars of five year paper. And they had 10 billion worth of potential purchasers. There is such a huge amount of money looking for a home. Everyone in the world (mid May 2002) is still plunking their money in the US dollar. If the dollar starts to slide and people want to repatriate their money we have big problems. Our interest rates will sky rocket in an attempt to get money that our debt demands.

COOK Report: And if interest rates go up the telecom crash accelerates?

p. 88 Just One Part of a Multipart Revolution

Googin: There are a lot of reasons why people are working on this. First of all corporate spending on IT will stop until this works. There are like six simultaneous equations that I talk about. Servers are going from multimillion dollar single system image servers to commodity "blade' servers. Instead of paying 1.5 million you will pay $50,000 for the same computing power. All these are totally dramatic changes. The second change is from the PC to the handheld. Third is Software becomes Web Services. Fourth: The local area network merges into the wide area network. Fifth are the semiconductor changes I have just outlined for you in detail and sixth the storage goes from server attached to network attached. This new paradigm will first get us to intracompany. That is to say all the corporate applications will finally talk to each other.

There will be an intra-company stage where people will say I control my standards and applications. And then once each company figures it out, the next stage will be getting it to work between the companies. This is when it starts to hit the network and when you need solutions for the bandwidth problems. I think traffic flows will be extremely high because you will treat the telecom network like it were part of you computer back plane. It will be like another interprocessor communication. That is the ultimate. You won't care where your storage is. You won't care where your different processors are.

It's the Latency, p. 89

Full Article

David P. Reed: It's the latency, stupid. (I coined that phrase a few years ago).

To be clear, what I mean is that end user (task) latency is what drives the decsion to use a communications experience. "Connectivity" and "bandwidth" ar special cases of reducing latency.

Bandwidth reduces latency by reducing the time the network takes to deliver all the bits needed to all the places where the task needs to get done.

Connectivity (i.e. always on-ness and ubiquity) reduces latency because it reduces the time it takes for the user to get his/her task "into the computer communications system" by reducing the distance they have to travel and the time waiting for connections to be made.

ICANN, pp. 90 - 100

Full Article

p. 90 The clique that runs ICANN has lived for the past four years in its own little dream world of deception and outright lies. The result for anyone who pays attention has been the destruction of trust in any mechanism of internet governance on the part of most outside observers. On the inside of ICANN meanwhile we have the reality distortion field of Joe Sims, Stuart Lynn, Vint Cerf, Mike Roberts, Esther Dyson and the rest of the IBM, ISOC, WorldCom clique who established ICANN in the summer of 1998. We shall visit that distortion field in this essay. The view of the insiders has been compared by ICANN observer Brett Fausett to the paranoia of Captain Queeg in the Cain Mutiney Court Martial. Fausett concludes If the blistering, paranoid rhetoric coming from ICANN officials sounds familiar, then you're probably a movie buff. We've seen this picture before. We can only wonder whether the ending to the ICANN story will be the same as The Caine Mutiny's.
http://icann.blog.us/stories/2002/08/17/whoStoleIcannsStrawberries.html

pp. 92 - 93 We suspect that the venom that ICANN's highly paid corporate attorney spewed forth had been simmering for a little more than two weeks. Sim's strategy had been grievously slapped down in a Los Angeles court room on July 29.

In Karl Auerbach's lawsuit when both sides (ICANN and Karl) called for summary judgment in the spring of this year, both sides, in effect, said to the court that the facts were not in dispute and asked the court to render a judgment as to what should be required of ICANN under law as a California 'public benefit' corporation. On July 29 the judge ruled decisively on behalf of ICANN Director Auerbach and on August 1 distrust of ICANN flared for the first time in a major way on the IETF mail list. The apolitical engineers of the Internet, listening to Dave Crocker lie and Vint Cerf prevaricate, came out and said enough.

Here is what happened:

Crocker: [snip] 2. Note that Karl won the suit, but lost the war. He is now subject to the conditions that ICANN had originally wanted to apply for his access. The only difference is that now Karl is under a court order to conform to those rules.

Froomkin: This statement quoted above in which it is alleged that Karl Auerbach got only what he'd previously been offered, is flatly contradicted by the text of the judicial decision. It is, quite simply, utterly false. In fact, almost the reverse is true: what ICANN got is almost exactly what Karl offered them months ago (the only differences being he has to give 10 days notice to disclose instead of 7, and some documents must be inspected on the premises). The judge stated that in several material respects, what ICANN sought violated California law.

See the text of the decision for yourself at http://cryptome.org/auerbach-icann.htm See also http://www.icannwatch.org/article.php?sidà3 for a subsequent development.

One does not have to be a lawyer to understand what the following means: the court said ICANN's position "violates both section 6334 and Bylaws Article V, Section 21 because it deprives Auerbach of the inspection rights he has under law and imposes such unreasonable requirements as having to sign a confidentiality agreement and having to pursue burdensome review in any effort to enforce his inspection rights".
p. 94 And on August 2, **Froomkin** in response to Cerf: both parties requested summary judgment

Yes. And almost all of ICANN's was denied, and almost all of Karl's granted. Your point being?

[...]

Cerf: No, ICANN would have preferred to stick with the procedures that were established. However it seems appropriate to point out that the principal difference between the procedures adopted by ICANN's audit committee and the procedures recommended by the court differed primarily in the manner in which differences of opinion as to the releasability of material considered confidential by ICANN would be settled.

Froomkin: To the extent this is an accurate identification of the "principle" difference, it is also one where the difference between the two views is substantial, and ICANN had its head handed to it on a plate. At INET, I tried to warn you this would happen. You clearly had been advised otherwise. That was very bad advice. You were so bought into it that you wouldn't even take the time to listen to *why* I thought what I thought. That was not prudent. [Editor:

Cerf: ICANN recommended an internal procedure that, if not agreed by the director wishing to release them, would then be settled in court. Your proposal, that was recommended with one modification by the court, required ICANN to seek judicial remedy if given a 10 day warning, it disagreed with the director's proposal to release. In your original proposal you suggested a 7 day period.

Froomkin: This is what we in the legal profession would call a sop for ICANN. If your lawyers are telling you differently, I submit you are again badly advised.

Cerf: Karl, [the] restrictions are still

there. If a director wishes to release information held to be confidential by ICANN, ICANN has a 10 day period in which to seek judicial review and restraint. At least that is the way I understand the order.

Froomkin: True. More importantly, a director has --and has at all relevant times had -- an independent, personal, duty to act in his best understanding of what's best for the entity. That is one which the law doesn't allow him to delegate. Or sign away, as ICANN proposed.

Cerf: Karl, even under the procedures adopted by ICANN to accommodate full review while protecting confidential information allowed you and any director full access to corporate records. The court did NOT validate unlimited ability of any director to unilaterally release confidential information.

Froomkin: Nor did Karl assert that non-existent unlimited right to release anything. Rather he asserted his statutory right to exercise independent judgment, rather than having it illegally constrained by a procedure adopted in violation of both law and ICANN's by-laws. Doesn't that finding bother you? Illegal. Violation of by-laws. Doesn't it make you wonder whether other things you have been told by the same people who drafted this illegal document are within the by-laws might not be? Wouldn't it be prudent to worry about that now? Once bitten, twice shy?

p. 95 Another Froomkin Law Review Article:

Editor's Note: Michael Froomkin also published a citation to a preprint of a new law review article, FORM AND SUBSTANCE IN CYBERSPACE, on the IETF list. In this 30 page article from our point of view he does two important things. One he summarizes in crisp and clear language of about 500 words how the formation of ICANN was accomplished in a way that ought to outrage anyone who believes that governmental authority out to be accountable to the law and to those over whom it is exercised. Two he points out how even in a law review article Joe Sims and his co-author Cynthia Bauerly's "chief rhetorical tactics are obfuscation and confusion." In our opinion ICANN's foundations are so egregious that its only shield is "obfuscation and confusion." Keep your enemies off balance and it will be much more difficult for them to understand what you are really doing. For context the IETF exchange introducing Froomkin's "Form and Substance" paper follows.

On August 1 on the IETF list Joe Baptista wrote: "but failing that complete technical control of "." is in the hands of the Internet's end users. That's where the power in "." resides.

Froomkin: I translate this to mean that (1) we can point to any root we want (true, but not real relevant for Joe Clueless User) and/or (2) the root servers could choose to mirror something other than the DOC root (true in theory but not real likely in practice, cf. discussion of root server operators in http://personal.law.miami.edu/~froomkin/articles/formandsubstance.pdf

ICANN to Do Crony Re-delegation of .org?

Meanwhile we have another perfect example of giving "government's power to delegate public functions, like the power to regulate, to unsupervised private groups." It looks like ISOC will be given the royal right to sell .org names. We quote in their entirety two brief articles. http://www.icannwatch.org/article.php?sid=927

Some ICANN Travesties for September 2002

ICANN launched a complaint against Verisign for having faulty woho is data. http://www.washingtonpost.com/wp-dyn/articles/A33395-2002Sep3.html
In addition to its complaint, ICANN today announced the establishment of a centralized online form (located at www.internic.net) that Internet users can use to report faulty Whois data. ICANN will also establish a tracking system to notify registrars of reported inaccuracies.

DoC Renews MoU

On September 23 John Paczkowski of "Good Morning Silicon Valley" writes:

"Commerce Dept. renews ICANN contract to mismanage Internet"
http://www.siliconvalley.com/mld/siliconvalley/business/columnists/gmsv/4134770.htm

Old Timer Gives Affadavit for RIIA, pp. 101 - 105

Full Article

Editor's Note: On August 15 we had the extremely unusual situation of Dave Farber coming out on behalf of the RIAA and against the position of the EFF. In an affidavit filed with the RIAA suit Dave took the position that it would not be unreasonable for the court to order the four backbones being sued to block access to a pirate file sharing site in mainland China. He was saying essentially that in defense of copy righted content a court could order changes in how the Internet is run. This unfortunate action was something new under the sun and greatly alarmed us. Fortunately the RIAAdecided to pick on Verizon and dropped its backbone suit on August 21. Verizon, we were pleased to see, came out squarely in defense of its subscriber privacy as it file a brief on behalf of Kazaa and against RIAA. We chronicle the course of these strange events below.

p. 104 ***COOK Report***: Why should any of us assume that a court will say that an ISP should block a music trading website but not block a website with political economic or social content that a deep pocket litigant doesn't like and therefore harrassingly sues for infringement under DMCA or Sony Bono? What about the hundreds of small political discussion sites, where individual posters frequently extract paragraphs or even whole articles from major newspapers so as to be able to share and discuss them with other posters. Should all these discussion sites be monitored by cyber-air marshalls, or simply shut down in advance to forestall the copying of copyrighted news?

Executive Summary

Whither the Policy Technology and Economics of the Interconnection of the Internet?

The collapse of the industry and of the price of bandwidth is bringing significant changes into the ways in which ISPs and the remnants of the Old Guard of Tier 1 backbones interconnect.

Some people who are affected have made some significant steps in using NetFlow data in developing tools that are being refined into what can function as bandwidth cost management systems. We identify several explorations being taken in this direction and explore what looks to be the most refined developed by Bill Woodcock with the assistance of Alex Tudor at Agilent Labs.

Bill has developed a philosophy of interconnection that appears to have a sound business model behind it. Bill's approach was developed from the point of view of a small ISP that needs to understand with as much precision as possible what it does cost to get its bandwidth delivered. His model says that ISPs that are multi-homed and have their own leased line customers need to peer as much and as cheaply as possible. They also need to have two reliable transit providers in case one fails. As long as their peering can cut over to transit if it fails, he points out that economics would seem to demand delivery of as much bandwidth by cheap peering as possible to cut down on the requirement for expensive transit bandwidth.

ISPs need to avoid local loop charges from their LECs and acquire their own back haul to an exchange for inexpensive peering and if possible a different exchange or exchanges for more reliable transit. In order to figure how to most cost effectively architect their networks they need to take and manipulate NetFlow samples of their traffic in order to identify potential new peers via a study of the traffic being delivered by their transit providers. If they have automated tools to take samples from appropriate points, they can over time get clear pictures of how their traffic is evolving through actual NetFlow path analysis.

But Woodcock's colleagues seem to agree that he has done something unique. He explains it in writing for the first time in this issue of the ***COOK Report***. Namely he does what he calls synthetic path analysis by tacking his actual path data and doing a series of "what if" transformations on that data. With the help of Alex Tudor from Agilent labs he explains using actual data from January 31 2002 how this synthetic analysis can be applied so that for the first time an ISP, by plugging circuit cost data into its modeling software, can know how much it really does cost to deliver its bits.

These ideas are new. While our experts agreed that perhaps100 ISPs may be doing some form of actual NetFlow data analysis, virtually no one except Woodcock had done the synthetic path analysis. Avi Freedman in his position as Chief Network Scientist at Akamai has had ample occasion to use network routing and DNS to figure out data flows. After studying Bill Woodcock's explanation found that he had evidence from his own related experience that indicated Bill's approach seemed valid. He points out that since 1999 he has been doing a what if analysis on "Akamai flows" similar to Woodcock's synthetic path analysis on router flows.

Our 50,000 word eight-week-long discussion involving 25 different people contains a quite interesting dialog between the Avi and Bill as they compare their approaches to the problem and conclude that the ideas appear to be valid. However, we must also point out that Bill's synthetic path analysis is not meant to be the sole criterion on which to base peering and transit decisions. Once they have identified potentially good peers, ISPs will find that factors of geography and costs of interconnection at various exchanges may become decisive factors in making their final decisions.

Although the largest carriers generally prohibit their technical people from participating in this kind of discussion, we were fortunate to get participation from large representatives of both the cable modem and DLS worlds (namely Adelphia and SBC). At the most general level these larger players seem to acknowledge the validity of Woodcock's ideas. However, one things get specific, they maintain that differences in the sizes of their networks prevent them from placing too much faith in low cost cooperative peering. They seem intent on joining and perhaps replacing the seven Tier 1 backbones.

As prices for transit bandwidth have crashed in about two years from as much as $800 per megabit-per-second per month to in some cases less than $100, the assumption that one should save transit costs by peering as much as possible has become muddied. Our contributors had a great deal to say about both the costs of transit bandwidth and city-to-city OC pipes of various sizes. Carriers are extraordinarily hungry for new lambda sales.

Farooq Hussain offers an essay that explains the international mess that he believes peering and transit has become due to the Tier 1 oligopoly. Although the view one gets of traffic depends on the places within the network where one measures, from some viewpoints according to Adlex and other tools it would seem that while UUNET, ATT, Level 3 and Sprint are holding firm in Tier 1, Genuity Qwest and Cable and Wireless may be slipping seriously in overall traffic rankings.

We reprint Bill Norton's comments on the death of ATM exchanges from NANOG. Our contributors explain why the price differentials aren't quite as crisp as Equinix would make them out to be.

Googin Interview on .13 Micron

We publish a mid May interview with Roxane Googin in which she explains

why the market arrival of .13 micron technology will bring PC economics and costs to switches. Nortel and Lucent look unlikely to rise and the LECs switching base will belong on the junk pile sooner than anticipated. The Googinization of the LECs continues - http://www.sbc.com/press_room/1,5932,31,00.html?query=20257. The LECs of course are intent to keep the secret bottled up and whine for regulatory relief that if delivered would only prolong the crash.

ICANN's Pathology

We summarize ICANN's latest egregious behavior from Vint Cerf's trying to spin the loss in the Auerbach lawsuit to Joe Sim's temper tantrum against Michael Froomkin. Froomkin has an outstanding new law review article called Form and Substance and written in no small part to lay bare Sim's obfuscations. From the predetermined award of .org to ISOC – sorry Carl Malamud, you don't count in this new commercial internet – to the regional routing registries public rebuke of Cerf and Lynn ICANN's misbehavior marches on. The DoC renewed its MoU this time for a single year. Bret Fausett's blog grows better and Milton Mueller author of a new dispassionate study of ICANN and the root finally throws up his hands in disgust on BWG. One of the most amazing things about the ICANN pathology is how, in the midst of the barrage of news about corporate malfeasance in Enron, WorldCom, Qwest, Tyco and others, the collection of figureheads who have agreed to serve as board members can continue to go on on a daily basis in blissful assumed ignorance that they have any responsibilities under California law.

We can only hope that Karl Auerbach who HAS taken his duties seriously while his fellow Board members have shirked theirs, will give some meaningful completion to his law suit by making it very public and very clear how this group of people who have assumed they are accountable to no one has in reality behaved.

RIAA goes Beserk

And Dave Farber unaccountably issues an affidavit on behalf of the recoding industry.

The COOK Report on Internet
COOK Network Consultants
431 Greenway Ave
Ewing. NJ 08618, USA

Empowering the Customer - $385 Now available. http://cookreport.com/empowering.shtml

Subscription Rates

Choice of either ascii or Adobe Acrobat (PDF) format 1. Individual; College or University Department; or Library; or Small Corporation - $295 2. Corporate - (revenues $10 to 200 million a year) - $375 3. Large Corporate- Revenues of $200 million to $2 billion per year - $475 4. Very Large Corporate- Revenues of more than $2 billion per year - $575

Site License: The right to distribute ascii and PDF via email to all employees of corporation. 5. Small corporate: $450 6. Corporate: $700 7. Large Corporate: $950 8. Very Large Corporate: $1200 . Site License Distribution via intranet web site $400 a year additional. See www.cookreport.com for more detail

Gordon Cook, President
COOK Network Consultants
431 Greenway Ave
Ewing, NJ 08618, USA
Telephone & fax (609) 882-2572
Internet: cook@cookreport.com

How to use interactive features of this PDF. Click here with hand tool.