The COOK Report on Internet

© Cook Network Consultants

A Content Versus Carrier Peering Battle in High Stakes BBN vs. Exodus Dispute

Web Farm Based Business Model Is Weak & May Lack Incentive to Engineer TCP Congestion

John Curran Discusses Best Exit & Paid Peering

Vol. VII, No. 7, October 1998

ISSN: 1071-6327

An Overview

In August a peering battle of a new and seemingly quite serious nature errupted between CRL, AboveNet, and Exodus on the one hand and BBN (GTE Internetworking) on the other. The dispute went public on August 10, when Randy Barrett wrote in Inter@ctive Week: "GTE Internetworking will drop free Internet traffic exchanges with Exodus Communications Inc. early next month and has sent similar notices to AboveNet Communications Inc. and several other Web-hosting providers."

The major difference in this peering dispute from earlier ones is that it focuses on a new content provider versus backbone carrier business model. In doing so it pits big web farms with minimal backbones against the major backbones -- in contrast to the earlier opposition of the smaller against the larger backbones. Many Internet engineers argue that the issue of access to content either has triumphed or is on the verge of triumphing over the issue of the need for symmetry in the exchange of backbone traffic. This view is expressed most clearly in an exchange with John Levine at the end of Part Two of this article. We believe however that it is decisively refuted in the interview with John Curran, CTO of GTE Internetworking that forms Part Three of this article.

A lot hinges on the business model assumptions under lying and motivating the actions of the opposing sides. The assumptions are neither widely known nor generally understood. Because, for several years, peering agreements have been clothed in non disclosures, the only time when much is ascertainable about the dynamics involved and business case assumptions underlying the agreements is when disputes arise. By keeping their cards too close to their chests the leading parties have been unable to educate the rest of the players as to their evaluation of the stakes involved.

As part of the fall out from the current dispute – the most serious since the move against its peers by UUNET in May 1997 -- it is to be hoped that positions and agreements will become more public. The stakes are growing. In contrast to past complaints of relatively small private companies, this time the grieved party is a publicly held company with a positive balance sheet of nearly $60 million. In addition, the FCC has also raised the possibility of becoming involved. As far as Exodus goes, as we point out below, our reading of the company's S1 and its most recent 10q reports leads us to conclude that it may well be over reliant on no cost peering with the backbones to which it is now pumping vast amounts of asymmetrical data flows. When Exodus and BBN settle their differences, the fall out from that settlement may give an even clearer picture of the puzzle.

While not discussing any of the specifics of the Exodus complaints, Curran draws a very reasonable analysis of the general peering issues involved. The asymmetry in traffic exchange between a web hosting firm and BBN, an asymmetry which varies but can be as great as 15 to one, destroys the assumption on which the Internet has so far been built namely that both sender and receiver must contribute to support traffic costs. The imposition of a "receiver pays" model for Internet traffic will lead to major major problems. Therefore it is incumbent for both sides to experiment to find a way to return to some kind of a sender receiver balance in figuring out how to pay for traffic from the world wide web.

We hope that, as a result of these disputes, the attention of all parties will now turn to ways to get more constructive solutions to these problems. And maybe, just maybe, their outcomes may be less subject to nondisclosure agreements which, in our opinion, have the very strong downside of keeping too many people in ignorance and raising levels of fear, uncertainty and doubt.

The Details of the Dispute

In his August 10th article Barret concluded: "Both Exodus and AboveNet were notified by letter that they would no longer be considered peers, which will effectively cut off the two companies from customers on GTEI's network, unless they pay to have data delivered. "We've been trying to be very flexible [with Exodus], but our traditional peering arrangement isn't one that works [for some large hosting providers]," said John Curran, chief technical officer for GTEI, which was formerly known as BBN Planet. The central problem is asymmetry of traffic between GTEI and the hosting companies, Curran said. Exodus pumps many times more bits to GTEI than are sent the other way. AboveNet would not comment. Exodus officials were surprised and angered that GTEI chose to drop peering and said the pain won't be unilateral. "It appears that the result of this action will harm BBN and its customers substantially more than it will harm our customers," said Exodus President Ellen Hancock in an Aug. 5

On the Inside:

BBN, Exodus & Peering pp. 1 - 11

IPv6 Survey pp. 12 - 16

BBN Engineering pp. 17 - 21

IANA not Open pp. 21-22, 24

letter to customers."

One of the events fueling the controversy was the posting by Net Access Corporation (NACNet) of Denville New Jersey of a scanned copy of a July 9th letter from BBN to one of its web hosting peers. (The name and address of the recipient were blacked out.) see http://www.nac.net/gte1.gif and http://www.nac.net/gte2.gif.

July 9th
viaFederal Express
Inside address crossed out

Dear xxxxx, GTE internetworking ("GTE") appreciates xxxx participation in the experimental interconnection, of our networks utilizing MED's and more specific routes. We have been evaluating this method of interconnection based on the information we have gained in the course of this experiment. [Editor's Note: According to David Woodgate of Telstra: "The Multi-Exit Discriminator, or MED, is a numerical metric that is used to indicate the preferred route to an autonomous system when there are multiple connections between two autonomous systems. When there are two routes announced for the same network number from the same AS, the route with the lower-value MED indicates the preferred route."]

There is a group of providers that source large amounts of traffic for GTE destinations -- larger than the public interconnects can support -- but that do not themselves transit significant GTE-provided traffic to their customers. This presents a number of challenges for a backbone provider, one being the imbalance of costs involved, another evolving our network in the face of large traffic fluctuations and increases from this group providers -- another large cost in time and network resources.

It was our intention in this experiment to determine if a scalable solution could be found for handling the distribution of highly asymmetric traffic flows in a manner which could become a standard aspect of our peering relationships.

Based on the experience to date we feel that using MED's and more specific routes to distribute traffic load will not lead to a scalable solution for addressing asymmetric traffic flows. As we have informed you previously, transit is an alternative to peering in the event that this experiment does not work out to GTE's satisfaction I would like to thank xxxx for their participation in this experiment as it has given us a basis for evaluating the relative merits of the possible approaches.

Accordingly GTE will provide 60 days from receipt of this letter until termination of the MOU and peering agreements between the parties is affected. A trouble ticket will be sent to xxx prior to any changes to remove interconnection between our networks. With the large amounts of traffic between the parties and xxxx desire to significantly grow that traffic load, GTE strongly recommends xxxx pursue transit services from a backbone provider who has significant direct interconnection capacity with GTE. Among those providers we interconnect with MCI, UUNET, and Sprint currently have the most capacity in the areas of your large data centers. Should you desire I can also put you in touch with our sales force if you want to investigate our transit services. Please feel free to call me at xxxx with any questions about this matter. Sincerely, xxxx Sr. Network Architect.

The Letter Leaks

Note that approximately a month passed from the receipt of BBN's letter until it surfaced on the web site of NAC Net, a North Jersey ISP. Note also the remark about "large data centers" that are critical to Exodus's strategy. AboveNet and CRL are much smaller than Exodus, and as far as we can determine have no "large data centers" spread around the country. Based on the vociferous nature of Exodus's public complaints, we conclude first that the stakes in resolving the CRL and AboveNet disputes appear to be much lower than the stakes with Exodus and second that it is highly likely that Exodus leaked its copy of BBN's termination letter which eventually found its way to NACNet.

To have this peering dispute burst open in this way is rather unusual. The industry standard has become one of cloaking these agreements in formally binding non disclosure letters signed by both parties. We may ask if there were no non disclosures in this case why Exodus would wait for thirty days to start fighting? We also wonder how it is that Exodus, a major hosting provider with significant requirements for peering due to its customers, had only a transitory agreement for peering with BBN? We wonder what was their fallback plan? We imagine that their customers would like to know when their other peering agreements will expire and what their likelihood of renewal is? The letter refers to an "experimental interconnection." By definition an experiment in peering has no assurance of permanence. Finally while in subsequent days of controversy on public mail lists, BBN's "greed" in trying to force Exodus to pay for maintaining a connection was roundly attacked, readers might also note that the letter suggests that Exodus seek transit from someone with good interconnection with BBN and not from BBN itself.

On August 17 after BBN had been excoriated by Exodus and other outraged smaller ISPs on the NANOG List for the entire week, we wrote the following in answer to a comment by:

John Levine: Exodus has claimed several times on NANOG, that although they peer with many providers, they don't buy transit from anyone, and if GTE won't accept their packets as a peer, they're not going to get their packets through anyone else.

The Exodus Business Model

COOK Report: I am sorry but I am really surprised at the paucity of research done by those who are bashing BBN. People really ought to read
http://www.edgar-online.com/quotecom/gdoc/?choice=2-885584&nad=0

There they will find Exodus stating in January 98 to the SEC that: "While the Company began operations in 1992, it did not offer server hosting services until 1995 and did not open its first dedicated Internet Data Center until August 1996, at which time it refocused its business strategy on providing Internet system and network management solutions for enterprises' mission-critical Internet operations. As a result, the Company's business model is still in an emerging state."

COOK Report: Translation it became a web hosting company and as the out put of its large content providers increased to the point where a two-to-one imbalance in traffic between BBN and Exodus is likely more like 10 packets or more into BBN for every one packet in the other direction.

Levine: Since Exodus hosts many of the largest and most popular web farms on the net such as Geocities, a lot of us expect that BBN's customers won't be pleased when they find that they've lost access to many of their favorite web sites.

COOK Report: John, this has been a point of contention on the internet at least since 94, but please show me where the consensus has been reached that web hosting companies need to be interconnected to the Internet for free. One IPO for 69 million and one bond issue for 200 million later Exodus has a very healthy net worth. Exodus started out as a backbone service. It migrated into web hosting. From the point of view of exodus, as Exodus admits to the Securities and Exchange commission above, this meant a new business model. Exodus may well have been fully peered with the big five back in 95 or even the big 10 or 15. Peering criteria were different in those days. Presumably their peering with BBN dates from that time frame ..... a time frame when traffic into and out of BBN might have differed by a factor of 2 or 3. But certainly not by a factor of ten to one which it is going to be, given the traffic asymmetry engendered by a web hosting enterprise today.

What those are saying here today is that a web hosting enterprise ought to be able to connect to the Internet for free. Remain a peer of everyone forever and not have to pay anyone for transit or for peering. If you let one class of business send traffic into (connect to) the Internet for free where do you stop? The ISPs on this list who are complaining very likely do some web hosting themselves. They make money at doing this and an expense that they bear is the portion of their upstream bandwidth that every megabyte sent out by their web servers consumes. Now if they pay, and Exodus doesn't they are subsidizing Exodus's competition against them. Why should they do that?

Later on the **Exodus S1** states (p.10): "In particular, the Company is dependent on WorldCom and certain other telecommunications providers for its backbone capacity and is therefore dependent on such companies to maintain the operational integrity of its backbone."

COOK Report: Exodus talks here about its layer two transport. Those looking to evaluate its strength should realize that in purchasing transport from a company that is on the verge of proving almost half of the Internet backbone services in existence today, Exodus has a critical portion of its infrastructure under the control of a competitor. Looking at the purchase by Exodus in late July 98 of transport from Qwest, leads us to conclude that Exodus has undoubtedly realized this.

S1: In addition, the Company relies on a number of public peering interconnections and private peering interconnections to deliver its services. . . . . Many of the operators of the private peering interconnections are competitors of the Company. Currently, the Company does not pay a fee for many of these interconnections, and if these organizations were to begin to charge the Company for utilizing these interconnections, or, in the cases where the Company currently pays a fee, to increase the pricing associated with utilizing these networks, the Company may be required to identify alternative methods through which it can distribute its customers' content."

COOK Report: What is not clear here is how many of the top 5, or 10 or 15 Exodus has private peerings with and how many it PAYS for peering or transit. It is clear from the above that Exodus PAYS - that it is NOT entirely transit free - despite the claims made on its behalf by various contributors to this discussion.

[**Editor's Note**: As we found out from the fallout to our questioning, it appears that what Exodus is paying for in some instances is peering rather than transit. Paid peering delivers its traffic less broadly than paid transit and presumably costs less although precise figures are impossible to come by since the details of paid peering agreements are covered by non disclosure agreements.]

UUNET seems to have gotten rid of most of its peers. When Exodus is presumably having to buy multiple DS3 per month connectivity from UUNET, should BBN subsidize both UUNET and EXODUS by giving exodus free connectivity forever?

[**Editor's Note**: we received the following assertion from a credible source: "Worldcom is paying UUNET for Exodus's connection because Exodus threatened to take their leased-line business somewhere else unless they got free connectivity." We also note that in the spring of 1997 when UUNET tried to terminate its peering agreements, some of the affected backbones put pressure on WorldCom by threatening to take their WorldCom leased line purchase elsewhere. We sent our source's assertion to Robert Bowman Exodus' Director of Network Engineering asking for comment. Rather than a simple "no comment," he bristled: "You're a sad excuse for a journalist. I can't talk about our relationships with other providers without their expressed written consent. And even if I had it, do you think I would talk to you about it?"]

Someone has to pay for the backbone infrastructure eaten up by web flows. Who benefits from web transactions has been a matter of controversy all the way back to the interview we did with Vint Cerf on peering in the fall of 1994. We still do have any clear answers but the industry is certainly not about to embrace a business model where content providers can send at no cost. They could do so but that would up the bills of web users and would be an invitation to an open season for spaming.

Levine: BBN's argument about the imbalance in number of packets makes no sense to me. It's as though my grocery store claimed that since the groceries I bought weighed 20 pounds, while the check I gave them in exchange weighed a fraction of an ounce, they're getting the short end of the deal. Networks peer because the two sides both get similar value from the connection, not because they get the same number of packets.

COOK Report: No, not the same number of packets, but at least there generally is some degree of proportion. I would like to know how Exodus can tell the SEC in January 98 that only some of its connections are free peering while others are purchased, and now, in August, whine that BBN is screwing it because *ONLY* BBN is breaking the peering relationship since Exodus is a customer of NO ONE else. If Exodus had gone from paid transit [or paid peering – editor] to free peering with one or more companies since the beginning of this year, one would think that Exodus's Ellen Hancock would have figured out a way to let the world know.

Of course Ellen ran networking for IBM in the late 80s and early 90s when IBM was left standing on the platform as the Internet train left the station. Then she was CTO at Apple while that company went deeper into hell. Now she's CEO of a company that made a 69 million dollar IPO in March, issued 200 million in bonds in July, and has just bought its OWN fiber national back bone from Qwest for an undisclosed sum. Now, when BBN stops giving it a free ride, (read multiple hundreds of thousands worth of free interconnections per year) whines and snivels over the whole Internet. Gee: why am I not impressed with Ellen's latest leadership?

Exodus' idea of how to do PR puzzles me. But I am even more puzzled that no one seems who complains on these lists seems to think to read its 10Qs or S1. If BBN and Exodus were peers as far back as 1995, the conditions of THAT arrangement have TOTALLY changed. For BBN to pretend that nothing had changed and continued to give Exodus free connectivity would be questionable management. Besides, as Exodus's latest 10q shows, it can afford to buy transit. Its assets exceed its liabilities by $57 million dollars. It has also just leased its own backbone on Qwest fiber for an undisclosed amount.

Part Two: Technical Considerations in Evaluating the Interconnect Burden Between Backbones & Web Hosting Companies

[**Editor's Note:** While we received a good deal of grief for our defense of BBN, Sean Doran provided useful technical input, pointing out that the imbalance would be in bytes not in packets.]

COOK Report (on com-priv on August 17): Oops. Thank you Sean. My point was data to BBN will be much larger than data from BBN to Exodus. Would it be correct to assume that the bandwidth imbalance would be on the order of 16 to 1?

Doran: It would be speculation, of course, and based on a guess of a mean packet size in the Exodus->GTE direction of about 512 bytes, and two such packets per 64-byte ACK in the GTE->Exodus direction.

If the packet mix has more large packets, then the imbalance in bits-per-time-unit would be larger. If the packet mix has more small packets, or there are more ACKs for some reason, then the bits-per-time-unit would be smaller. However, I think that 16:1 is probably the right order of magnitude.

COOK Report: Are you saying that unless Exodus's ingress bandwidth into BBN were kept large enough, such that packet loss and retransmissions were avoided, that the effect on performance of the interconnect could be severely magnified?

Doran: Well, that's obvious, but not what I was saying. Congestion leads to drops, but a packet dropped before arriving at BBN's border router hurts Exodus rather than BBN.

Exodus Impact on BBN More Than Just Matter of Bandwidth

(However, if there are system effects, such as there are with many types of shared switches -- most notably DEC Gigaswitches -- then packets being dropped in the exchange fabric does have a chance of interfering with other GTE-bound traffic. That is, if Exodus's transmissions were to cause a queue in a switch facing an interface which GTE uses to accept traffic from many parties, those transmissions in effect worsen those other people's connectivity with GTE.)

It is congestion after the first GTE router - *ANYWHERE after, including congestion in customer networks* -- that is the problem. Loss after that point causes the same data to occupy a part of GTE's infrastructure twice. TCP's congestion avoidance is wonderful at minimizing the number of retransmissions needed to move data from sender to receiver. A more aggressive set of algorithms in effect robs GTE (and everyone else right up to the point where the congestion occurs) of otherwise useful bandwidth.

COOK Report: Suppose Exodus were a customer and not a peer of BBN. Would these technical performance issues be any different?

Doran: If they are doing proper congestion avoidance, no. Either the customer connection would be not-full, or Exodus is seeing drops which rob itself of capacity. (And is also bad since the transmitting hosts have to send more data than necessary). If they are not doing proper congestion avoidance, then there is exactly the same problem no matter how money flows.

Note that poor congestion avoidance can happen even with adequate TCP implementations, because of how many short TCP flows behave in aggregate. ("Being nibbled to death by a herd of mice" is how Jon Crowcroft put it). A key point about web servers is that they are sources for large numbers of short TCP flows. An interesting draft to look at is draft-ietf-tcpimpl-prob-04.txt which indicates that there are some badly behaved TCP implementations that are commonly deployed without the users' knowledge.

Then **Michael Dillon** wrote: So if packets have value, then Exodus is sending 10 times more value to BBN than they are receiving.

Doran: No. Remember that TCP transmission rates are clocked by the reception of ACKs. The transmission of more packets than that which would be transmitted by a host implementing the congestion avoidance algorithms in draft-ietf-tcpimpl-cong-control-00.txt does not add value, but SUBTRACTS it. A gross imbalance, in other words, removes bandwidth from the target network, in substantially the same way as a flood-based denial of service attack does.

Michael Dillon: I should have said, if a packet's value is determined by the byte count...

Doran: Well it's not. A huge packet that is dropped before it arrives safely at the destination, and is subsequently retransmitted, has reduced the amount of bandwidth available along the path between source and destination. In other words... A packet that ultimately is not delivered has negative value. A large packet that is ultimately not delivered has large negative value.

Dillon: And as you point out, not only are BBN customers requesting those packet flows but BBN's routers are setting the actual rate of flow.

Doran: No. Remember that packet loss can happen anywhere in the network. BBN can have constant 0-length queues, but if its customers have queues, BBN still suffers when packets are retransmitted. (As do the customers).

Dillon: This would be a case of a mismanaged network

Doran: No, it is a case of how short TCP flows behave in aggregate, and how TCP implementations have subtle bugs that lead to less than ideal congestion avoidance.

Dillon: And I don't recall hearing BBN make the claim that these peers are mismanaging their networks or peering interconnects.

Doran: BBN and its supporters, you might have noticed, have been very quiet. They are demonstrating the principle of not transmitting into a too-busy mailing list. The same is not true of BBN's detractors, and Exodus's supporters, the sets of which overlap to some degree.

Dillon: I am assuming that this is a situation where the number of packets sent into BBN is substantially the number required to carry the traffic requested by BBN customers.

Doran: I dunno; it's of course impossible to tell without someone in the path kindly providing a TCP trace which shows the number of retransmissions from a source behind Exodus to a destination behind BBN. Ideally that number approaches zero, but as BBN has many receivers reachable through it behind interfaces with very small buffers (i.e., less than tens of buffers per interface), it seems unlikely that it approaches it all that closely.

Dillon: I think that you are getting confused here. The packets enter BBN's network because the customer requested them. If the customer can't handle the packet flow then that is their problem.

Web Traffic Impact on TCP Congestion Avoidance

Doran: Woah!

1. TCP congestion avoidance is THE stability mechanism for the Internet (source: Van Jacobson and company, echoed readily by many researchers and some clever "doers")

2. TCP congestion avoidance is done by the transmitter.

3. When a client requests data to be sent via a TCP connection, it expects the server to transmit according to TCP rules, which mandate that the sender must be congestion-avoiding.

4. If the source does not do congestion-avoidance, it is not the receiver's "their problem", it is a packet flood.

5. Be conservative in what you send is a philosophy that is absolutely required to avoid classical congestion collapse. Sending packets that are likely to be discarded due to congestion *anywhere* or due to a receive window overrun at the recipient itself is not conservative.

6. Sending traffic much faster than the maximum rate at which one has *discovered* the network can accept is not conservative, _even if the requestor

has *asked* for the traffic_.

7. Being unconservative in either point (5) or point (6) is TCP-unfriendly behavior, and deserving of opprobrium, and probably also deserving of "Denningeresque" hyperbole, and even possibly deserving of actual consequences.

Dillon: If BBN doesn't like their customers requesting large packet flows then they should deal with the issue with their customer, not cut off their peers.

Doran: The problem is not large packets; the problem is sending more unacknowledged data than the maximum of the discovered congestion window or the advertised receive window. Large packets are in fact generally considered good. Returning to [my earlier statement that] A packet that ultimately is not delivered has negative value. A large packet that is ultimately not delivered has large negative value. I should have added: A large number of packets that is ultimately not delivered has large negative value.

Dillon: If BBN doesn't like their customers requesting large flows of packets then it should deal with the issue with their customers, not cut off their peers.

Doran: I don't understand why you don't see very clearly that the problem is not with the volume of data but the rate at which it arrives relative to the rate at which any bottleneck along the path can accept new traffic. BBN customers include several organizations that certainly, as part of their function, do regular bulk-transfers of data -- huge flows that dwarf any WWW transaction. The problem is not in the size of the flow, or in the size of the packet, but in transmitting packets when *any* part of the path from source to destination is congested. Moreover, in the absence of congestion control in routers, it is absolutely the sender's responsibility to ensure that it avoids contributing to congestion anywhere.

Russ Nelson: Maybe I'm on the wrong mailing lists, but this is the first that I've heard of this. Are you saying that someone's TCP stack is broken, and that BBN is ceasing to peer with them because of it? If that's the case, then all the flaming is wasted hot air -- fix the frigging stack!

Doran: Stacks should be fixed. Please do not follow me down the path of tying the general problem of broken TCP stacks at big web farms in general to the GTE/Exodus peering. I raised it as a possibility rather than suggesting that in general web farms might be able to eliminate a whole category of concern by being deliberately TCP-friendly, which probably would have been more effective. Sorry about the confusion.

[**Editor**: In answer to a statement by Rahul Desai -] Talking *generally* about peerings, if one is connected to a large source of TCP-unfriendly web traffic, the traffic is not particularly unwanted, but rather arriving in a destructive fashion.

Since, unfortunately, Cisco's RED is not as stable as I'd like before seeing it deployed widely -- despite, I might add, some efforts also on the part of John Hawkinson (who as you may know is at BBN Planet, and pretty clever at kicking routers) -- there are ways that one can rearrange one's topology to lessen the effects somewhat.

If the effects cause problems, and one cannot use RED to randomly drop segments from TCP flows to match the input rate at any congested interface to that interface's output rate, and one has none of the other proposed mechanisms for dealing with TCP-unfriendly traffic, then one could, in principle, use someone else's (congested) network to absorb the overload for you. It's a gross and funny hack. Don't take it too seriously or believe that this is what GTE is actually doing.

Before Russ Nelson says, "well, duh, fix the routers", I should point out that that is what we are trying to do, as I said during my presentation at the last NANOG. Note also that even in a general sense, the TCP-unfriendliness may be purely unintentional on the part of the large traffic source network, and may not easily be fixable or even fully understood by them.

Meanwhile on NANOG on August 18, **Curt Howland** commented: The 'Net gives some measure of universal connectivity because it was driven at its root by engineering. Who on NANOG, with any real world experience in networking, denies that maximum open peering benefits everyone?

Paul Vixie: Well, I don't know if I'm qualified by that measure. At the time I built networks, I'd built at least one of the largest anywhere. But they were all small compared to today's second tiers, and I'm not building networks any more.

Maximum Open Peering – Who Benefits?

But I deny your assertion. Maximum open peering benefits some people more than others -- specifically the ones who get to charge the most money yet who pay the least in infrastructure upkeep. A web hosting company doing shortest-exit (no matter how many peering points they were at or how much private peering they had) would be an example.

Remember as you puzzle your way through this issue that peering is only mutually beneficial if the number of bits (not packets) sent by each side is in the same order of magnitude. If the O(mag)'s differ then the costs/benefits are one-sided, and the side who is underwriting wide area transportation costs for people who aren't paying it money is going to get bent about it. [**Editor's note**: although Vixie doesn't name names, his statement describes the Exodus BBN dispute from a point of view friendly to BBN.

Anyone with a strong enough constitution to check the archives on this matter will find that Sean and I had a raging battle here about this very topic back in 1993 or so. What interests me on this particular night is my memory of asking Vince "so what about gatekeeper.dec.com? why should I have to pay to transmit the FTP archives?" Vince didn't answer because the obvious answer ("gatekeeper should be charging money so it can cover its costs and the costs of the folks who carry those bits") was one I was not at that time ready to hear or understand.

Doran: ftp.gatekeeper.com, like all content hosting sites, attracts users to one location, when it could instead attract users to the same content in multiple locations. This is inefficient.

Many archive sites direct people to mirrors topologically closer to them, and as more actual content owners purchase duplicate sites in places like Europe (http://www.europe.cnn.com, http://www.yahoo.{se, .co.uk, .ca}) both for the technical gains and also as a means of strengthening one's brand globally. This is much more efficient. This should be carried on.

It is in carriers' interests to encourage the trend too, and it is probably a good idea to work out means of hosting content more locally than data centers in California, as a means of increasing network efficiency.

It is in content-owners' interests to encourage the trend also, as ultimately their brand is the one that is hurt by unreliability. Yes, it might hurt carrier X's business when carrier X cannot get to Popular Content Site, but some fraction of carrier X's customers will go away thinking, "performance to Popular Content Site sucks! Popular Company sucks!"

Moreover, as the capacity market distortions are sorted out in many places, it almost certainly will be cheaper for content owners to spread the work of distribution around, so that the traffic stays as local as possible.

Returning to com-priv: **Rahul Dhesi** wrote: Are Exodus and other web providers using poorly designed implementations of TCP and causing the problem? Or are their TCPs ok but the

unfriendliness is caused by the nature of TCP?

Doran: Let's depersonalize it. Don't think in terms of Exodus vs BBN. Think of it in very general terms. A large source of short TCP flows sends traffic along several branches of a tree rooted at the source. If at any point in the tree the rate of the combined traffic from the sender in question and all other senders exceeds the interface rate at that point, queues will form. Individual TCP flows will back off as the queues grow and in particular when they fill (as then drops happen).

The problem is that the large source in question is sending multiple parallel flows along the same part of the tree, and two things will tend to happen:(1) as one flow is signaled (by drops and delays) to slow down, another flow begins speeding up and (2) as one flow is signaled (by drops and delays) to slow down, another flow starts up for the first time

Therefore, while the individual flows will share bandwidth adequately, there is no relief for the overloaded interface. Worse, any other flow (from another source, for example) will have backed off and will keep backing off because it is trying to avoid adding to congestion. Note that this can happen anywhere on the tree, including at the last hop (say, a dialup modem).

Desai: If the latter then perhaps it's BBN's customers who are responsible for the bad data, by sending the SYNs in the first place and causing the unfriendly traffic to be sent back.

Doran: You could argue that, as some popular browsers open multiple connections to the server, this behavior (assuming a *perfect* TCP implementation on the sender side) is BBN's customers' fault, and perhaps you have a point. However, you should look at Scott Huddle's message, which states that a sender is always free not to send. Moreover, Paul Vixie has developed some experience in building web sites which do not reply to all four (or more) simultaneous TCP sessions, and other folks have done sterling work on having multiple parallel TCP flows share the same congestion window. Either approach ought to be taken by a source of "mice" (small TCP transactions).

However, if you want to argue that any destination ought to be the one to protect itself, then the appropriate technology is intercepting caches. Even with zero caching gain, one can do useful things to arriving traffic, in exactly the same fashion as the sender ought to have been doing in the first place.

The bright side is that an intercepting cache that *does* have caching gain -- including that acquired by deliberately making content less dynamic than the source wants it to be -- will tend to reduce the amount of traffic (and perhaps countable views) emanating from the "bad" web farm, thus costing it money.

> Curran: if we enter a world where the senders don't really pay any incremental costs, you face some huge implications. You end up with a situation where, for example, a sender could decide to send you a large video image when you connect up to his web site. . . And I guess I am just a little bit concerned, if not from a business perspective, from a public policy perspective about the multimedia spam possibilities when senders are not paying any real incremental costs for sending more information.

Finally, there are a couple of techniques being studied as prototypes for congestion-sensitive traffic admission and for recognizing and penalizing traffic flows which do not back off correctly in the face of congestion, and this will help all users of the networks except the sources of traffic who charge their customers based on number of packets or number of bytes measured close to the source (rather than total number of correctly delivered bytes).

The "Value Equation?" Is Content Gaining the Upper Hand?

Finally on Tuesday August 18th John Levine answered our comment on compriv – herewith an excerpt: ***COOK Report***: John, this has been a point of contention on the internet at least since 94, but please show me where the consensus has been reached that web hosting companies need to be interconnected to the internet for free.

Levine: Doesn't matter in the least to me. But I expect that it matters quite a lot to BBN's customers who will be peeved to find that BBN no longer provides connections to many of their favorite web sites. I sure would be. Peering is based on equal value which in a sane world means that the two sides customers are equally eager to hear from each other. That's why Japanese backbones have had to pay for their own connections across the Pacific, Japanese users are very eager to get to US networks, but few Americans care about getting to Japanese networks. In the Japanese case, I believe the packet flows are about equal.

[**Editor's Note**: At this point, we took the discussion private with John and thank him for his permission to publish the results]

Levine: I don't see anyone arguing that BBN and Exodus haven't been getting proportional value.

COOK Report: Whoa! The value point can be argued until the cows come home.

Levine: Well, sure, it's up to the two peers to decide if they're getting reasonable value. But BBN's letter didn't say anything about value, it only talked about counting packets.

COOK Report: I meant proportional traffic flow. With a web based asymmetry for every t-1 of bandwidth that BBN sends Exodus, BBN had to take about half a t-3 in data in exchange. Why should BBN GIVE such bandwidth to Exodus? I very much doubt that Sidgemore is giving it away.

Levine: If his customers want it, why shouldn't he? They're paying for Internet access, Exodus has content that they want. You could as well say that UUNET or BBN should pay Exodus for access to the content that lets them sign up those customers.

I can't see any reason why the number of packets passed through a peering point should be symmetrical if the kinds of users on the two sides happen to be different. It makes no more sense than saying my car should weigh the same on the way home from the grocery store than it did on the way in.

COOK Report: It seems to me what you are arguing for is a world where you network provider must interconnect to content providers at no cost to the content provider.

Levine: Well, if that's what it takes for the network providers to sign up and retain customers, yeah.

COOK Report: where does a model like this stop? I don't want to have to pay for the expense of my ISP having to give free connections to every content provider in the world. spammers would love that arrangement i'd think

Levine: Your ISP will pay for the connections it needs to get its customers what they want. Which would you prefer, a $20 connection to an ISP that can get you to Geocities, Excite and Infoseek, or a $15 connection to an ISP that can't? Spam is a red herring here, it's not content, it's anti-content, and they should be paying ISPs to accept it.

COOK Report: Let's try looking at it in the following way. You and I own back-

bones and we agree to peer. Fine.... but then *YOU* turn to web hosting and suddenly you are sending me 16 bytes for every byte I send you am I obligated to increase your bandwidth into me by 16 fold for free?

Levine: Depends. You ask your dialup customers whether they'd mind if they lost access to my web sites. I'll ask my content providers and advertisers how important it is to retain access to your dialup customers. Then we can figure out who's willing to pay what share of the interconnection. Note that the number of bytes passed isn't part of the equation here. At the moment, it looks to me like the dialup users are more eager to have access to every web site than the web sites are to have access to every dialup user, so the money flows in the direction from the dialups toward the web farms. Maybe there'll be a time that will change, but not yet,

Peering in the US developed in a homogeneous Internet, where all the NSPs had similar kinds of customers and a similar mix of data sources and sinks, so everyone was equally eager to connect to everyone else. But look at the connections between the US and the rest of the world. For years, international networks have paid the full cost of expensive trans-oceanic links because their customers demanded access to US network resources, whereas most US users didn't care whether they could FTP stuff from Australia or Europe or not. (A few did, but not enough to affect the pricing equation.) That's the model we're moving to, with specialized networks bringing different amounts of value to different kinds and numbers of users, and anyone who thinks he can measure value by counting bytes is going to get run over with a steamroller.

Probably the closest parallel is cable TV. Some cable channels full of infomercials pay CATV systems to accept their feed, others like HBO charge those same CATV systems for the privilege of carrying their feed, and others like broadcast channels neither pay nor charge. Better get used to it.

The Many "Faces" of Peering

[**Editor's Note**: peering – always hard to track because of the non disclosures involved has taken on a variety of forms which in the absence of clear definition can make intelligent discussion of the topic even more difficult. The following contribution by IBM's Sean Butler to com-priv on August 17th was as good a summary as we have seen in a long time.]

Butler: I've done a bit of reading over the past several days, and want to make sure I understand the possible solutions to the peering dilemmas today's Internet faces.... Here are the ones I have found, but please let me know if there are others, or if I have missed any of the issues with the ones I list. This is a very short review so only the major issues are listed.... It is assumed peering means free exchange of yours and your customer routes/traffic. I am not advocating any of these possibilities!

I. Open Peering

This is where everyone peers with everyone else and its all free. I think everyone agrees that in this scenario, local and/or regional ISP's get a free ride on the backbones of the National ISP's... Or, if 'free-ride' is a bad word to use, at least the large ISP must long-haul traffic across its BB to the exchange point.

II. "No rules" peering.

This is basically where we are today. ISP's are free to peer with any other, publicly or privately, as long as both agree to the arrangement. The issue's are that most ISP's don't disclose what the requirements are to peer with them, and that there is much debate over what "peer" means as far as the relative value of each ISP. This "value" is not defined and different people have different ideas about it. If I send you 10M worth of content since you sent me 1M worth of requests for that content, who has more value?

(Of course this is a gross over simplification as we can all tell from the volumes of posts on this in the past week over BBN's announcement to pull peering from web-centric ISP's, but I'm just trying to cover the basics!)

III. Settlement / reciprocal billing based peering

David Holub's postal example explains this well... When you send a letter from the US to Mexico, you just use US postage, so Mexico is basically delivering the letter without payment. The US sends 100 M letters to Mexico, and Mexico sends 70 M letters to the US. The country that delivers less mail pays the country that has delivered more to make to 'even.'

Of course, for the Internet, what counts as 'mail?' Is it packets or bandwidth or what? And for those that currently 'peer,' i.e. no money changes hands, there will be great debate about who should be paying whom. Also, what happens when there is a flood based DOS attack. Do you separate that out and not count it? (Hey, if you did have to pay for sending BW to the other ISP, NOC's would certainly fix attacks faster!)

The router technology may not quite be there to do this type of accounting, although it is not far off. The real issue is how do you define value...

IV. Common carrier status based peering

Another idea pushed by Holub... Basically, if you are an ISP and you have common carrier status, you have the right and obligation to provide 'universal service' or 'universal connectivity.' I.e. other ISP's (with/without?) the same status must peer with you...

In this case, the telco's are required to terminate calls to their customers from other telco's, and vice versa.

The common fear listed hear is that as a common carrier an ISP may have to pay access charges to connect to the LECs CO. But does this price outweigh having to pay for transit?

V. Regulation

The government steps in and regulates peering... If ISP X stops peering with Y, so X's customers can no longer reach Y, there is no 'universal reachability' which the Internet was formed around, and though the NSF turned the network over to private enterprises, it was under the impression that universal connections would not go away... So, the government steps in and regulates peering to ensure the Internet is not segmented in anyway.

Or, a private independent body does the same thing...

VI. Brokered Private Peering

As proposed by Savvis, Exodus, Electric Lightwave, and Williams communication.... (See *Boardwatch*, May 98 cover story, for the white paper.) This proposal covers a bit more than 'peering' in the traditional sense as it defines a layer 2 infrastructure and Service Level Agreement's, but ignoring that for now...

Basically defines two main classifications. Primary classification is national, regional, or local. Secondary classification is business- customer oriented, consumer dial, or web-centric. Based on those classifications, rules state that 'true' peers must peer with each other. Non-true peers can apply for true-peer status in other classifications.

A non-true peer can either 1. negotiate for free peering, 2. pay for peering, or 3. buy transit from another network...

Part Three: John Curran Discusses Some of the Policy Issues of Peering

Editor's Note: John Curran is chief technical officer of GTE internetworking and should need no further introduction to the readers of the COOK Report. We interviewed him on Friday August 21. What follows is we believe the most detailed and intelligently articulated exposition of the issues involved with peering and the direction in which the industry needs to push these issues that we have seen to date.

Curran: While circumstances prevent me from talking about issues still outstanding with any particular peer, I will cover how we see connectivity in the Internet working today; what happens when we extrapolate that model going forward with more peering agreements and involving more hosting companies; cover some of the problems that this creates; and then discuss some possible alternatives.

Let's look at the evolution of peering in the Internet and, for your readership, which is quite sophisticated, let's hit only the high points. Major backbones right now exchange traffic via peering. In the case the very high traffic flows, the exchanges are done through private peerings (private interconnects as op-

posed to public interconnects). In general other Internet Service Providers connect to a company that is either one of these backbones or is someone buying transit from one of the major backbones.

Pricing Model: Pay to Send and Pay to Receive

In the end, if you take a look at how pricing is done on the Internet, most folks both pay to send traffic and pay to receive traffic. Look at any one network provider out there and consider a customer of that provider on one coast and a customer of that same provider on the other coast. If you look at your utilization, you will find it both the sender of the bits and the recipient of the bits are being charged for their Internet utilization -- in other words for their traffic. This is the case on almost any usage based plan that you can find out there. This is not to say that everyone is usage based in their charging. There are some flat rate plans out there. But, in general, people are paying both send and to receive. I have yet to see anyone advertise a "receive only" Internet connection or a "send only" Internet connection. Of course, I'm refering to dedicated Internet services purchased by corporations and ISP's, not consumer Internet pricing. Consumer Internet pricing has far more factors and cost of transmission may not be the dominant one.

COOK Report: So you are saying that usage-based plans do tend to have something in them that says we measure both your incoming and your outgoing traffic?

Curran: Yes, from what I've seen in the industry, this is quite common. Sometimes it's the average of those and sometimes it's the greater of the two. You have to work with individual companies. If I were to poll people on the Internet, I would find that almost every company connected to the Internet has the assumption that, the more traffic it sends, the more it pays.

COOK Report: Pay more for sending more absolutely. In our impression what is a bit less common is paying more for receiving more.

Curran: I think that corporations connected to the Internet with corporate networks know that their employees are mostly pulling traffic off the Internet. It is quite common for them to expect pay for that access.

COOK Report: And, in effect, you are telling us that, with these new pricing plans, this is what happens because they are measured on what they receive as well as what they send?

Curran: That's quite common. Now, if you take a look at this practice, you find that there are some interesting ways in which you can bill everyone. You could actually bill of both the sender and the receiver the full freight cost of getting that traffic from one end of the country to the other. Now this would be really good if both were on your network, and you could bill both. But this is not a practice that I support. I would actually like to have the sender and receiver each pay about half of the costs associated with their action so that when you add the two together, you end up recovering the full cost. You should not recover your cost by a factor of two by billing each for the full amount. And, if you look network pricing today, I think this is the way a lot of it is done. People assume that they are being compensated by both the sender and receiver -- that both are paying some part of the overall share if both are on your network. Now this is the one network model.

When you peer with other networks, things become more interesting. Let's take the case where the sender is on your peer network on the West Coast and receiver is on your network on the East Coast. The sender sends traffic to his ISP peers with you and hands it off very quickly because of the glory of shortest exit routing on the Internet. As a result, you have to carry it from the West Coast to the East Coast. Now I use the example of West the East, but this would be equally true if it were Chicago to Dallas or Seattle to Tampa. The fact is that the receiver is paying the bulk of the cost associated with getting that traffic onto its network because they receive it very close to the sender and bring it very close to the recipient.

Now when we undertook peering, we all knew this to be the case. And we all recognize that traffic which you receive from a peer you don't get paid to carry.

COOK Report: But the traffic goes in both directions and, in the end, it will balance out?

Symmetry of Traffic Flow

Curran: Let me go into that a bit. The fact that I am receiving traffic for which I am not getting paid doesn't bother me because I'm taking equal amount of traffic -- approximately equal -- and I am sending across my peering connection to the other side. This is traffic from my direct customers, which I am being paid for, and which, thanks to shortest exit, I don't have to carry very far. Now to the extent that I have a peering connection and I'm receiving quite a bit of traffic that I'm not being paid for by the sender -- I'm looking at the downside. But there's an upside. I'm taking quite a bit of traffic that I am being paid for, and I'm able to hand it off without having the carry it to the other side of the country. So there is a balance there. As a result, the fact that the sender of traffic is a customer of the peer coming in from another network is not a problem. Because while this situation results in costs to me without any revenue, the good news is that, on your end, it also provides a reciprocal opportunity to have revenue with very little cost.

So this is how peering has worked for some time. We're really all very well aware of these circumstances. It is nothing new to say that from time to time providers get together and talk about traffic going between their networks because they have to engineer our networks properly in order to make possible such inter network exchange. But, in this situation, you don't want to get too far out of balance because, if you do, you end up with a situation where your costs don't match your revenues and with a business model that doesn't scale.

Now, in the peering relationships that we've maintained we've always advocated that some form of symmetry should be present. We are not talking about the exact number of bytes in both directions. What we're talking about is an order of magnitude. We realize that traffic varies from network to network, but realizing this, believe that networks should strive to present equal value to each other.

COOK Report: Is it your opinion that GTE Internetworking's interest in maintaining this symmetry has generally been shared by other large players in the industry?

Curran: Absolutely. And in fact I will say that the most important thing to recognize here is that the reason you would like to have a peering relationships which have roughly corresponding value is that, by doing this, you can continue to add more peering interconnections without getting into other alternatives like settlements. . .

COOK Report: Or business models that are so unbalanced they don't work?

Curran: Exactly. We've seen that trying to maintain this equality among peers lets us actually stave off doing a bunch of things in the Internet that haven't been done before -- like having to seriously consider other economic models such settlements. In order to keep the network running the way this today, we felt it to be in our interest to try to advocate that peers have balanced traffic. And by balanced traffic we mean traffic that is not wildly asymmetric. We do not want to be in a situation where the traffic in one direction is four

or five times the traffic in the other. Two to one may be acceptable. More than that is generally not. We have some cases where we send more to a peer and other cases where a peer sends more to us. As long as it falls roughly within a bracket of two-to-one in one direction or the other, we say the both parties are working toward the same goals.

COOK Report: Is your goal of two-to-one in the same ballpark as the peering criteria of the rest of the industry?

Curran: Let's put it this way. We all know that when numbers get far outside of two-to-one we have concern. I think that there are some folks who have some concern when numbers get more than 30 or 40 percent in either direction. We happen to believe that there is a lot of room for leeway here, and that two to one, either way, is close enough for now.

COOK Report: And your position on this in the industry is on the liberal side?

Curran: It is. We have a balanced set of customers. These include very large corporate connections. We have downstream ISPs. We have a lot of dial-up customers, but we also have significant web hosting. We find that we are quite balanced with almost all of our peers. However, if we find an ISP that has very little web hosting, we find ourselves sending them slightly more than we receive. Likewise with ISPs that are primarily web hosting firms and have no way to terminate traffic and no access customers, we find ourselves on the other side. In general for our 50 plus peers we find that we are pretty much in the middle of the line. For the vast majority of these peers this is not a problem. People realize that need to maintain some form of equality. We don't focus on a ratio of one to one but rather on a situation where the symmetry of traffic is not wildly out of line.

COOK Report: A very interesting part of this model is that it strikes us that it will work for a backbone that is considerably smaller than that of BBN. Why? Because it has been a commonly held perception that the in order to peer with someone, you have to be just about as large as the network with which you would like peer.

Curran: Well I will state very clearly that we apply these criteria, no matter whether the connection is public or private. In fact we believe that this is the type of model that works even when we turn up peering with a new company at one or more of the public interconnects. We see it as an important goal to allow peering and to allow people to get connected to the network. This is not an attempt to raise barriers. Rather it is a strategy that says we are willing to recognize symmetrical value even though it be with the smallest of backbones.

Now I guess I do need to point out that the fact that we do try to maintain roughly symmetric traffic in both directions, in and of itself, isn't very difficult. By this we mean that companies recognize it and it is not a big surprise. But over the last year or so we have seen a case or two when sometimes traffic gets out of line. We have seen folks who send significantly more traffic than they receive and these often turn out to be the web hosting firms.

Web Hosting Breaks Symmetry

It is interesting for me to hear that this shouldn't be a concern for us because our customers are requesting this data. I actually have had the opportunity within the last few weeks to discuss with quite a few of our customers what they expect. And all the customers to whom we've spoken have agreed that they want the content that is represented by such sites. But they expect that the content producers will be sharing in the costs of getting that traffic to them.

What is interesting is that a lot of organizations we've spoken to have actually spoken to a lot of organizations having the content. These content producers also believe that they should be sharing in the cost of getting the content to its recipients. We believe it makes sense. The recipients believe it makes sense. The people who produce content believe it makes sense. In fact many of them believe that they are paying for it right now.

So the interesting fact is that we now need to make sure because of the nuances of shortest exit routing on the Internet, when you actually put that transaction together, it turns out that the recipient's network gets all the cost.

I want to make clear. We are not innovators here. Even the firms that participate in hosting have recognized this fact. About 5 months ago, a group of them got together and wrote a paper on the topic. The group included Savvis, Exodus, Winstar and some others. They put together a paper on what they called Brokered Private Peering. They recognize that the current model, as put together, because of shortest exit routing, doesn't scale. It doesn't provide for the distribution of costs necessary for us to continue to provision infrastructure. Therefore they put together an alternative.

Now they talked about the problem and found that the problem was clear. When you took on traffic with shortest exit routing, and for which you were not being compensated, or for which you are not receiving an equal value in your ability to terminate traffic in the other direction, you let yourself in for a bad deal. The fact of the matter is that these providers recognized the issue at hand. Consequently, they put together a proposal. Now their proposal happens to parallel something that we've been experimenting with. This is called the longest exit, or best exit routing.

Longest Exit Routing

Here is the way it works. Take an entity with a lot of outbound traffic -- a web hosting provider for example -- and see what happens, if, instead of handing it off to us at the first possible moment it carries it to the destination pop or facility and then hands it off to us. There are number of reasons why these companies might be willing to do this. Many of them have their own networks. Or they want to develop their own networks. Why would you hand off traffic to another company if you were building your own network infrastructure? Why would you not want to use that infrastructure? Likewise, doing it this way, let's you control the quality, because you are providing your customers a service. The longer the traffic stays on your network, the more control you have over the quality of that service.

In the discussions that we have had with a number of hosting companies, these companies have expressed interest in this solution. We actually did undertake an experiment to work with longest exit routing. And while there are some possibilities here, there are also some technical and business issues. For example, it is really not enough to carry the traffic only part way, you have to make sure you have a relationship where both parties agree what longest exit means, how far you are carrying the traffic, what the hand off is, and how you allocate costs. This is one solution. And we think it is one that needs exploration and further refinement.

COOK Report: Presumably the July 9th BBN letter posted on the NAC net web site has something to do with your experiments in longest exit routing?

Curran: We have not commented on the letter posted on NAC Net web site. Let me continue. Now, although longest exit routing could be a solution to some of these problems, it does have some problems of its own that I think we need to worry about. The first one is that taking those sorts of routes presents some scaling issues because those routes do not aggregate well They have some significant routing implications as well as requirements for interconnections that may not scale. So I'm not saying it is a solved problem. It is one with real technical issues and real business issues, but nevertheless, it is one that is worthy of further exploration.

COOK Report: And you feel this to be the case regardless of the Exodus situation?

Curran: We will not comment on particular peering situations. I can say that we have done this once and it didn't work out very well. So we are looking at the situation and will be examining multiple models on which we could move forward but I can't say when and with whom we would do that exploration. We will continue to explore these problems and search for solutions.

Paid Peering

Once solution that is, -- and for many providers it is eminently practicable -- is to look it the resulting costs. As I said, our receivers pay for some of this traffic but they also want to sending companies to pay for some. Now in this situation, looking at the incremental costs that we bear, we could say rather than your adding the infrastructure, we will do it, but we expect to be compensated for that by some form of paid peering, or settlement based peering. This idea has actually been very well received by a number of companies and it is actually not that difficult to do. To be fair, what we would want to do is to first say that, if the traffic is in balance, no one pays anyone else. With this model in looking for payment we would only look at traffic above the two for one ratio acceptable for no cost peering.

COOK Report: And only then would payment kick in?

Curran: And payment would only be for that traffic. We have to make this as reasonable as possible. We have found a great reception for this idea and have worked out a number of such agreements.

COOK Report: Do you mean that you have a number of such agreements in effect?

Curran: All I can say is that I am giving you examples of some of the kinds of discussions that we are having in the industry. The factor the matter is that this model also allows people to think about where they want to interconnect without worrying about which providers have which traffic destinations where. Rather than having to build a mesh of interconnects between your network in every major backbone, you connect where you want although you may incur a cost for doing so. But if you want 2.1 megabits of traffic and you need to hand that traffic off in Washington D.C. rather than in Chicago, then pay a fee and you can do so. You don't have to worry about what would likely be the greater expense of running circuits up to Chicago just to carry that traffic there. If you hand it off to the other provider and pay for the Delta, it will probably work out, both in engineering and in terms of performance, to drive the other person's backbone up to Chicago and pay the small incremental amount. It would certainly seem to be a win for everyone and therefore this is a model that we certainly think is worth exploring. I can also see a hybrid variation of this where people connect in some places in the country and pay to get their traffic to the remaining spots.

Receiver Pays Internet

The only model that I don't see as viable is one which a few folks are advocating -- one that I will affectionately refer to as: "receiver pays Internet." This says basically: ignore the problem. It says it that your customers are trying to get at this so they should pay. It also says that the sender network bears no cost and that, by choosing shortest exit, they can dump it onto the receiving network at the closest interconnect. The problem with this is if we enter a world where the senders don't really pay any incremental costs, you face some huge implications. You end up with a situation where, for example, a sender could decide to send you a video image when you connect up to his web site. After all why should he not send you a video image? He doesn't pay for it, you do. With all the multimedia development going on and with the ability of people getting on with cable modems and ADSL, the ability of sender's to pump data is going to be enormous. And I guess I am just a little bit concerned, if not from a business perspective, from a public policy perspective about the multimedia spam possibilities when senders are not paying any real incremental costs for sending more information.

Now I think there are a lot of cases where, if I connect to web site, I might want to see the 15 second highlight of a particular sports play. But I also think that there is a reasonable cost trade-off that the sender and receiver will make in deciding whether or not it's worthwhile to send out. But on the other hand if the receiver pays everything, what you do when you connect to the site and I send you the four minute corporate overview video? I simply believe that decision-making will become very skewed in a receivers pay everything Internet.

COOK Report: And there is a difference in sending the 15 second sports play as a default for everyone who connects and sending it only on demand.

Curran: Sure. The fact of the matter is that people claim that the receiver wants all the traffic he gets and that's why he should pay. All I can say is that there are some models where it is very clear that the receiver does want the traffic. But remember also that not all traffic falls into this mode. There is traffic that you get from advertising banners. In other cases you may get to a research site where the receiver has already paid for the privilege or receiving information. Why should he also pay for the transit? I would maintain the companies that send the data should pay for the transit and, if they believe that their receivers should bear the full cost, then they should charge the receivers as well. I will also note that the companies doing the sending get the benefits not only other receiver but also the advertising vendor. In a situation like this there is value being exchanged multiple levels. All that I am advocating is that at the network level we have in economic model that makes sense and we make sure that someone who sends traffic and someone who receives traffic both pay their fair share of the burden. That's not to say that, at a higher level, that a web hosting site cannot recover its costs. In fact I think a lot of them believe that they do get charged for it now and that they do recover their costs. In any case, at each level of the infrastructure let's try to keep a balanced economic model.

COOK Report: well, just to dot the "i" and cross the "t" of all this, is it correct to assume, as John Levine says earlier in this article, that it is becoming more and more commonly accepted that network providers must interconnect with content providers at no cost to the content provider? In fact his opinion seems to be if that's what it takes for network providers to sign up and retain customers that's okay with him. What you think of that?

Curran: Although I haven't had a chance to speak to Mr. Levine on that opinion, is point of view seems to me to be simplistic. Anyone expressing that point must realize that the Internet is not made up of just content providers or just access providers or just network providers. You need to have in economic model that makes sense. We can set up an industry where only the receivers pay which seems to be what he is advocating. And an industry in which the content is perceived to be all valuable. But I will tell you in that world, as a receiver, I will not take any ads because I have to pay for them. It's just like a paid cable TV channel does not get any ads because the receivers drive the boat.

We don't have that world right now. We provide network services and transmission services. Those services are independent of the content. I don't look at the packets and say: is this is a packet from a web farm. I must have a customer getting value from it. I can tell whether it is from a corporate web site or a government web site. It's just a packet. In order for people involved in Internet services to continue to scale the network, we really need to operate without looking at packets and by making sure that costs are recovered. There's no magic bullet here. There's an assumption for example that the customers you

sign up actually do desire access to the web sites. Some do, but some don't. And while some want access to the web sites, those web site providers also want access to those customers. I do not believe that it is the job of network provider to get themselves between the end-users and the content providers. I think the content providers work out their own exchange of value with some having web sites that you must pay to access. Now indeed there may be some customers very upset when they can't get to them. But they will be upset with the fact that while we are willing to pay some of those costs, and content company is willing to pay for some of those costs, the hosting company does not in all cases have a fair allocation of costs back to the network?

I think when you ask: who was going to be absent? You will find that there will be a lot of people upset. We honestly believe that most of the parties here are willing to pay their fair share. It is just a question to making sure that you have an allocation model that works. There is no company out there that is just an access company and there is no company out there that is just a content company. Advocating that we should sort ourselves into those roles in order to meet an arbitrarily defined business model seems to need to be an overly simplistic view of the world.

COOK Report: Would be fair to assume that anyone in the ten or fifteen largest Internet Service Providers holds a point of view on this subject that is markedly different from the one you have just expressed?

Curran: Well I think, if you look particularly at most of the people involved in major Internet backbone service, you will find that they believe that traffic is something you receive and send to customers; that you get compensated for that by customers; and that traffic has a cost that needs to be recovered. I don't think any of us are looking at the bytes to see exactly who is getting the higher value in the exchanges that are going on.

It is true that on the very fringes of the network you will find some very simple business models. There are some folks who buy transit to the network and sell access to dial up customers. There are certain number of dial-up customers and certain cost to their transit. Now if the world were made up of just those companies, people who sell access to dial up, and just web hosting firms, you could probably wedge everyone into the business model that receiver pays. Luckily the Internet is a far more diverse place. We have access companies and hosting companies and backbones and regional networks and corporate networks. We have many sites to originate content on a paid basis and many sides originate content on a free basis. We have recovery of costs through advertising. The world of the Internet has many more transactions than the access company that pays to get at the content site.

I do think this diversity is a good thing. Simply because we have an aberration of Internet routing -- namely shortest exit routing causing problems with a particular cost recovery model -- I do not think that this aberration in Internet cost recovery should cause to structure the entire industry into the overly simplistic view of receiver pays.

How Do We Solve Our Problems?

COOK Report: Can we conclude by asking where you think we need to go from here?

Curran: We need to continue to look at the options, specifically longest-exit-routing and settlement-based peering. With respect to longest-exit-routing, we need to examine the technical .. We need to examine technical possibilities for scaling this. We need to look at the business model that needs to be behind it. To be honest and fair I will say that longest exit routing may actually be unfair in the opposite direction. Longest exit routing puts the cost disproportionately on the sender. And I'm not saying that that is a fair world either.

We need to ascertain the if the right model to follow it is not the one where the costs get allocated between the receiver and the sender by splitting the total cost of both sets bits that they sent. In fact longest exit may not actually be a fair allocation method if we believe that the cost of the transmission should be borne equally by requestor and sender. So we need to experiment. And I must say that people who have glib short answers to this may not have actually thought about the full implications for fairness on either side.

Now you are not going to want to maintain a mesh between all those who originate content and all those network backbones. (For example ten geographically dispersed interconnects from every one of 20 different content companies to every one of 15 different backgrounds.) This would certainly create a routing nightmare that we don't want to live with. There may be some cases where you would want fewer interconnects in the context of an industry model that allows people, in places where they don't want to build infrastructure, to instead pay fair cost to get their traffic there. So I think that we're going to see both some work in longest exit routing and some cases in paid peering to get traffic to regions where a content provider doesn't have infrastructure will also emerge. I would hope that we can key both paid peering and experiments in longest exit routing down to a minimum because it is very good idea to keep the network as simple as possible.

COOK Report: Presumably you will go ahead and explore these directions. Do up you see anything that the industry, as a whole can learn from this?

Curran: I think that this set of events over the last few months has actually served a valuable role in bringing attention to this issue. It is an issue that we as an industry haven't paid much attention to. It is true that there have been discussions going on, but they have been fairly low key. The nature of discussions in general, because they are done under NDA, means that it is quite common for them not to appear in the press. That is not the case this time around. If there is an upside to this, it is that it has raised the level of this issue of everyone's mind and we can be thankful for that.

Sources of Information About IPv6 - see p. 16 below

Editor's Note: these urls and titles of Internet drafts were sent to us by Jim bound on July 24, 1998 and were updated by him on August 20th.

Compaq IPv6 page: http://www.digital.com/info/ipv6/ The 'white paper' by Jim Bound and Al Cinni hot linked to this page provides a good general comparison of IPv4 and IPv6.

6bone page: http://www.6bone.net/

IPng (IPv6) Page: http://playground.sun.com/pub/ipng/html/ipng-main.html A very useful collection of pointers to engineering documents and IPv6 implementations.

Useful Internet drafts:

Overall IPv6 Addressing Architecture: http://www.ietf.org/internet-drafts/draft-ietf-ipngwg-addr-arch-v2-07.txt

Global Unicast Aggregatable Addresses: http://www.ietf.org/internet-drafts/draft-ietf-ipngwg-unicast-aggr-05.txt

TLA Assignment Rules: http://www.ietf.org/internet-drafts/draft-ietf-ipngwg-tla-assignment-05.txt

Case for IPv6 from the IAB: http://www.ietf.org/internet-drafts/draft-ietf-iab-case-for-ipv6-02.txt

NAT Issues from the IAB: http://www.ietf.org/internet-drafts/draft-iab-nat-implications-01.txt

Also here are the drafts being worked on to avoid the existing NAT strategy:

draft-borella-aatn-dnat-00.txt

draft-tsirtsis-aatn-mech-00.txt

draft-montenegro-aatn-nar-00.txt

draft-ietf-ngtrans-assgn-ipv4-addrs-00.txt (this is the one I wrote I told you about)

draft-ietf-ngtrans-header-trans-02.txt rfc2356.txt for Firewalls where NAT is not needed by use of tunnels.

IPv6 to Begin Deployment Next Year
Limitations of NAT Technology in IPsec, IP Telephony and Mobile Computing to Force Corporate Moves
Number Allocation Is Hierarchical With Operational Improvements

Editor's Note: Jim Bound is the IPv6 technical leader for the Compaq (formerly Digital) Unix group. He went to the Amsterdam meeting of the IETF in July of 1993 to understand what was being proposed to solve the IPv4 address completion problem. He has been involved ever since in the development of IPv6. He also implements and architects IPv6 in Compaq-Digital Unix products. We interviewed Jim on July 23.

COOK Report: While it is evident that Brian Carpenter and others in Europe are interested moving rapidly ahead with IPv6 deployment, and while we have been told that the Japanese consumer electronics industry would like to have IPv6 deployment as soon as possible, the situation in the United States seems very much different. People say that we have such a huge IPv4 installed base that moving to IPv6 will be extremely costly and take a very long time. Furthermore, some suggest that the with CIDR and NAT boxes the IPv4 address problem has been solved and there's no longer any great urgency to transition from IPv4 to IPv6. How do you react to that as a summary of the current state of affairs?

Bound: I have heard similar reports from our business people in Asia. As far as the US is concerned, I am not so sure that the deployment of IPv6 depends on address space as much as it does on that the presentation of a solid business case. It is quite true that deployment of IPv6 will mandate a significant investment of resources from everyone doing it. I think that what people must do is look at the business case and get past the marketing and past Band-Aid solutions like NAT boxes.

COOK Report: Of course if you look at CIDR as a solution to IPv4 address shortage, you have the complaint that it pushes power upward toward the large ISPs by forcing the small ones to get their addresses from the larger ones. And a problem with NAT boxes is that they tend to break some protocols.

Incompatibilities of IPv6 and NAT

Bound: There is an ongoing effort in the IETF now to work on network address translation and document what this technology just won't do. Right now it won't support the IPsec protocol. NAT boxes also break H.323 for audio and video. The first problem is a show-stopper for anyone needing the security protection of IPsec and the second stops IP telephony applications. Any protocol that wants to use both addresses and mobility doesn't operate well in a NAT box environment. The mobility issue comes into play if you take a laptop outside of your corporation. The corporate gateway needs to put its address inside of packets destined for your laptop so that when you want to confirm back to your host your eligibility to receive them, you are unable to do so. The whole idea of a NAT box is that any addresses appearing in the header are going to be altered, so the mobile node (roaming) really does not know where their home agent is regarding an IP globally routable address.

Network address translation as a technology has quite a few shortcomings. So what is happening from my view of the world is that people are just trying to do basic things like Web lookup's on AlteVista. FTP will work as a long as the NAT box will save the address of where the connection has to come back to reach the client.

Therefore, the question that I think has to be asked: is what are people doing with the Internet? The bottom line is that, with a NAT box, if there is any data in the header that is required for end-to-end connectivity, it can be lost. The other problem is that if I have an address inside the NAT box there is no way to get to that address directly from the outside.

Now people would say that this is a security advantage. I think it really is not because the minute you let anyone in, whether it is through NAT or not, you have compromised your situation and the network node. The fact is that once there is a node on an intranet that is also on the Internet, you are compromised. What you need then is a security protocol. But by definition NAT breaks with IPsec which is the, I would think, emerging security protocol for the Internet.

IPsec

COOK Report: A year ago, we did an interview with Bob Moskovitz about IPsec. How has it evolved in the meantime?

Bound: First of all the core standards are very very much completed in my opinion. It is a very complex topic that people had to agree on. It has come together very well including the key management portion. There are two aspects of IPsec. The first is the gateway; the other aspect is the transport mode of IPsec. The gateway mode says that you go to a gateway where you provide IPsec and, in agreement with the other end, you tunnel and establish a security association. The tunnel would begin at the border of the autonomous system and could be the border of a private network. What Moskowitz is doing with the Automotive Network Exchange is a good example. So IPsec will be deployed at gateways until we get a lot of robust implementations. Everyone has made commitments to do it and are working on it. IPsec uses addresses as part of its security parameters index and association. When you start translating it breaks.

COOK Report: You mention IPsec implementations. How about a quick thumbnail sketch of the major implementations?

Bound: I believe Cisco has an implementation. Microsoft does. I believe IBM does. Based on the IPsec bake-offs. There are commitments for implementations from Compaq, Sun and others.

Now the principal of NAT breaks the principal of end-to-end connectivity -- namely that packets should go from origin to destination without being examined. Now let's take a common sense approach and say: if you are performing alterations to packets, you cannot possibly perform as well in that modality as you could if you did not examine or alter packets. If you look at the numbers of people using NAT, I believe that most do so because they are forced by circumstances to do so.

COOK Report: Take a company like Texas Instruments which has been through a massive renumbering of its IP networks and never wants to have to do that again. TI has been using in network address translation at its firewall. Would it therefore be a good candidate for transition to IPv6? How would this be decided?

Bound: First of all in IPv6, IPsec is mandatory. So every IPv6 system shipped into Texas Instruments will contain the IPv6 security extensions.

COOK Report: Might Texas Instruments be driven to this point of decision by

needing to be a certified trading partner of the Automotive Network Exchange?

Bound: That is quite true. But remember IPsec also works with IPv4 but it cannot work as well. The reason is that IPsec, if used with IPv4, has to have another option associated with the header. In IPv4 you have to say oh, when you get this packet, look at what they call the next protocol ID which will show you an encapsulation for IPsec that you'll have to parse. In IPv6 all of this is merely an extension of the header. It is inherent in IPv6.

We have efforts in the IETF designed to avoid network address translation. I have written a draft. Someone at 3Com has written a draft. Folks at Sun have put forth several drafts. Several of these are designed to create a mechanism which basically will come up with a way of assigning an address for someone who was in a private network so that when necessary they have a mechanism to speak securely and with a global address. Right now the working group is called AATN: Avoidance of Address Translation.

COOK Report: So if you move to IPv6 a you can avoid having to do this translation. What then is the reason for working on this AATN protocol nonetheless? Is it to give those companies with problems some kind of an interim solution?

Some Deployment Issues

Bound: That's a good question. Let me say that none of us have been able to totally predict how IPv6 will be deployed. One thing that we know you can't say to a customer is here's how you must transition from IPv4 to IPv6. Such a tactic couldn't work because every customer has a different need. For example let's say a company, XYZ, wants to transition one large division to IPv6. OK. We transition that division and life is good. But the first thing that happens is that division will need to talk to other departments that are still IPv4 within the same company. The priority then becomes to bring this communication about without network address translation. What we have discovered that when a new IPv6 node wants to talk to an old IPv4 node, we give that IPv6 node a globally routable IPv4 address just for the duration of the communication.

And let me add that switch vendors, router vendors, workstation and PC makers, when they ship IPv6, will not ship only IPv6. They will ship both IPv4 and IPv6, a dual stack. You will have both stacks to use. It will be an evolution from IPv4 to IPv6. It will not be like turning a light switch on.

COOK Report: Well in a router with a dual stack -- what happens when a packet hits such a router? Is there something in the software that looks at the packet header and says oh: IPv4 send it that way or oh: IPv6 send it the other way?

Bound: Yes. But we also defined an Ethernet type, and we can define a PPP type or a T1 type so that, when it comes right in at the data link level, you can see that it is an IPv6 packet.

COOK Report: So you know by be physical link the packet uses, what kind of packet it is?

Bound: Right. We defined it for Ethernet and we need to define it for other types of data links. Also there is a version number in the header. Consequently in our implementation we come right up through the stack and through the driver and see the header at which point say: "oh it's a v6 or a v4." Then we did some interesting things in v6 where we defined compatible v4 addresses and designed an architecture so that you can build what we call an IPv4 mapped address. All this gives us the ability to get an incoming packet's IPv6 destination. We have tested all of this quite satisfactorily at the University of NH where we have an IPv6 test bed with v4 <-> v6 interoperability tests every three months. The router vendors have gotten quite good at going back and forth between v4 and v6.

We are also looking for avoidance of translation for IPv4. In other words, if you are on a private network and you decide that you want to talk globally, you go off to your DHCP server where you get a temporary v4 address that enables you to talk v4 across the Internet and avoid NAT. So NAT is a necessary part of the evolution today for someone who is doing only minimal kinds of communications from a private network to an Internet.

What you pointed out about Texas Instruments is very relevant. Even if a company like that is hugely dependent on NAT, once it starts dealing with suppliers who need to use IPsec in their transactions with it, the security needs of the communications will render continued use of NAT impossible if you want pure end-to-end communications between nodes.

COOK Report: But while they will not be able to continue to use the NAT boxes, they will be able to avoid both renumbering and a forced march transition to IPv6?

Bound: Yes, but at a cost we have reiterated here.

COOK Report: How will the introduction of IPv6 in IPv6 routes and routing affect the use of CIDR and the question of the size of IPv4 routing tables at the default-less core of the global Internet? Can I tunnel an IPv6 session across the current backbones of MCI and Sprint from California to New York with out any of the routers on the two IPv4 backbones being aware of my IPv6 session or routing?

Bound: We actually do this today, it is real-time, with what we call the IPv6 backbone test network -- the 6 Bone, which operates globally.

COOK Report: But all the routers actually see are the IPv4 addresses?

Aggregators in IPv6 Architecture

Bound: Yes. But let me say a few words about the IPv6 architecture which, from the Internet perspective, is far superior to that of IPv4. We have defined the use a top-level aggregator that you can give to 10,000 ISPs and can extend to another 200,000 ISP's if need be.

IPv6 addressing is definitely going to change things. In IPv6 the only people who will be assigned top-level aggregators or top-level bits in the address blocks will be large backbone providers who route bits all over the world. So what it does is to take the CIDR concept and make it more robust and efficient by saying that, at that level, all they have to route on very quickly is really only the top order bits. Therefore, it has to be much faster. Beneath the top order bits is another set of bits called the next level aggregator. And then below that is a set of bits called the site level aggregator. ISPs assign to their down-streams bits from the appropriate aggregator level.

COOK Report: So one advantage of doing it this way is that each aggregation level would set borders that could not be permeated by route flaps developing within them?

Bound: Correct. Impact's from route flaps should be much reduced. Routing tables would become smaller.

COOK Report: Is there likely to be any kind of debate about the qualification's for ISPs being able to be assigned what level of aggregator and the right to hand out addresses from what level of aggregation? Suppose I'm growing in size and I think I should be eligible for top-level aggregation addresses and not second level?

Bound: There is a common sense rule set to determine that. Namely that when you have used up 90 percent of your next level aggregator space you can ask for top-level aggregator space. We have

proposed to the three IP number registries that there be a top-level aggregator. Every industry player who is routing bits will get a sub top-level aggregator referral that will give them 24 bits of next level aggregator space. Twenty-four bits means that they should not complain about running out of space anytime soon.

COOK Report: Are you saying then that in the IPv6 world allocations of addresses within one aggregation level will be defined as globally routable at the next higher level?

Automatic Renumbering in IPv6

Bound: Yes. But let me clarify. In IPv6 you get your address from your provider, unless you are a provider. And then you either go to IANA or one of the three registries. Now when you switch providers, assume that you are going to renumber. However, please note, that in IPv6 we have built-in automatic renumbering. IPv6 has built-in automatic renumbering from the autonomous system through the backbone and all the way down to the desktop. Renumbering from the routers to the end nodes has been tested and we find it works very nicely.

There is a draft out now on router renumbering. What this draft will do is establish the mechanisms for renumbering all router prefixes. Because we structured the addresses in pieces, the router has to only restructure the prefixes. In IPv6, when you boot your workstation or desktop machine, you have an IP address the moment boot process begins. We have something called a link local address so that every node on the local area network without talking to any human or to any server has an address. This will provide the very large benefits of dynamic renumbering and auto configuration.

COOK Report: Someone recently commented to us that IPv6 will really get underway when Bill Gates puts and IPv6 stacked in the release of Windows 2000. How do you react to a statement like that?

Bound: We have to face the reality that the market share for the average desktop and therefore for the average Internet user is Windows 95. And clearly IPv6 will not be pervasive at the desktop level on a worldwide basis without Windows 9X. However, this does not mean that IPv6 will not start very serious deployment before a new version of Windows is released with an IPv6 stack inside.

COOK Report: Then how about some scenarios by which it is expected to start deployment?

Voice over IP the Killer Application for IPv6?

Bound: Voice over IP is a prime killer application for IPv6. This is simply because IPv6 has a flow label which, on an intranet, can provide quality of service so that IP voice packets can be sent to a server or router that gives priority in forwarding them. Traffic bearing the flow label is given priority over traffic without it. In this context one also may deploy IPv6 servers even though they are surrounded by IPv4 nodes.

COOK Report: Let's assume then that I am a company like Merrill Lynch with offices throughout the United States and around the world and that I am buying Vocal Tech and Vienna Systems IP telephony gateways and I am really moving as much of my voice telephony traffic as possible off of the PSTN onto the Internet. I can do this now under IPv4. Why would I be really better off with IPv6?

Bound: There are two answers to your question. One is that if you have a dedicated voice over IPv4 gateway, and that all that gateway does is forward IP packets containing voice, this will work with IPv4. Now if you start putting other functions on that box like a web server or a database, you will have multiple kinds of traffic coming into the box. If you want the telephony traffic to have a higher priority than the other traffic, with IPv4, there is no elegant way to reserve resources. In IPv6 you have a flow label so that you can assign Quality of Service parameters to that voice call and you could even given one person more priority than another person.

COOK Report: In such an IPv6 set-up, if I am system administrator, could I assign some people a high voice call service priority but some of the same people lesser priority in calls to other locations and block some people from calling at all?

Bound: Yes, as long as we are talking about service on a single intranet. When you hit the public Internet, you get into a whole new scenario because there you must cross provider boundaries. But IPv6 will facilitate the signing of service level agreements (SLAs) between providers where each provider will undertake to provide a well-defined service level for the other's traffic. They will do this by undertaking to honor a specified service metric that is to be found in the flow label of the IPv6 packets. Without such a service level agreement, you have to go with what is called the type of service bit found in IPv4 or the traffic class in IPv6.

COOK Report: But going across service provider boundaries in IPv4 is the whole reason for the current differentiated services working group. Is what they're doing in the differentiated services working group really only good for an IPv4 world? Or can be carried over into the IPv6 transition?

IPv6 and Differentiated Services

Bound: The product of that particular working group can be good for IPv6 world. The bottom line is that you will get no more in IPv6 than you will get IPv4 because all they are doing is defining per hop behavior. One thing to be aware of is that the vendors make most of their money on the equipment for intranets. There RSVP and Quality of Service will work fine.

Note also that in the terms of service byte there is a bit that can be set and when the bit is set it will say go look in the flow label of IPv6 to get additional information on how to differentiate this service. But Brian and Kathy have asked me not to bring this up for the moment and I am willing to go along. I'm not sure why they are doing this but I think that it may be because they need to solve, right now, the very basic problem of what is the type of service given to these bits for IPv4 and for per hop behavior that will make the routing tables most efficient?

COOK Report: Does the amount of effort being given to differentiated service, right now -- and the effort seems to be substantial -- mean that everyone involved things that for the next year, two years, or three years there will be enough service sensitive traffic across provider boundaries to make all this trouble worthwhile? What is the life span, for example, of IPv4?

Bound: While we will have more and more IPv6, the life span answer to your question is at least another five years.

COOK Report: So even if we are talking differentiated services over lease the next five years, we are talking service is with the dollar value that falls at least into the multiple billion dollar range?

Bound: That is true. But, if we are talking IPv6, there is nothing to prevent several differentiated services vendors getting together and adding to their implementation IPv6 enhancements so that they can get service level agreements by means of the IPv6 flow label and not even bother with the IETF. In other words just get vendor consortia together to create the these IPv6 enhancements. All that needs to happen is that router maker "x" and router maker "y" and desktop vendor ABC all agree that, if this bit is turned on, we will honor looking in the flow label to see if

there are instructions there. This is a set of actions that does not require a standard.

COOK Report: So you are assuming that when the group comes together in this fashion, it will form a catalyst that will attract other members to that group?

Bound: Correct. Because the IETF moves way too slow now. There are so many people who, that not much work gets done despite the fact that many of the groups are very good. There are so many vendors there who are proposing different solutions that it takes a very longtime to sort it all out. In some cases it is almost better to wait until everyone is about done, and, if, at that point, it doesn't work, go do it another way. Of course if you have a certain set of bits that you don't want people to step on, that is when filibustering starts in the IETF.

COOK Report: And one reason it is so slow, is that the current mechanism just doesn't scale to the current size and number of players?

Bound: I would not say it's the mechanism. Rather, what I think needs to happen is that all groups should do what Brian Carpenter is doing with differentiated services is where all the groups need to go to. That is they have to stay on track.

COOK Report: So where does that leave us? For IPv6 adoption, are there big groups like the automotive network exchange? Are there big industry groups? Or big hardware platform or OS groups? You have some early adoption of IPv6 beginning: what are the variables that are going to affect the players involved in the adoption of the course that they take?

Bound: let's go back to IP telephony. With 32-bits a cellular phone company does not have enough space to support all the phone numbers with all the addresses that must be assigned to those phone numbers. So companies like Bell Atlantic, NYNEX, and Cellular One are going to have more subscribers than 32 bits will provide. So from that perspective, the people who will be doing the deploying include Vendor customers who absolutely have to have the connectivity. They cannot get it right now to the degree that they would like because we're still in the process of building out and the deploying all the basic parts as we finish up on the routing for the new network for IPv6.

But today there are parts of the world which simply cannot exist with 32-bits of address space. China for example cannot continue its existence with something based on only 32 bits. And you are simply not going to do NAT for all of China. NAT use in the US is viable for situations only of minimal complexity. NAT users cannot be doing mobile nor quality of service, nor IPsec. And then if I were to build a large wireless entity with say 20 million nodes operating, the first thing to be said if it is wireless, I have to have viable security as a part of such a network.

COOK Report: If you look at wireless network development, is it well known how much of it will be spread spectrum which is much less susceptible to any kind of interception?

Bound: Not by me. Although it is true that spread spectrum is relatively secure. But rather than the means of transmission let's focus only on the size of this network. If one goes to IANA and say I want 20 million IPv4 numbers, you are going to be in for something of a problem. The lesser amount of time that the industry is under such a restriction of numbers, the better. Now if no one has any of these needs NAT is fine if one can absorb its restrictions for the networking paradigm.

IPv6 Address Allocation

COOK Report: How do you then get to the point where you can begin allocating IPv6 addresses? If I am beginning some IPv6 applications, I must be using IPv6 addresses? If so, where do I get them?

Bound: We are developing procedures for allocation right now. I predict by March 1999 you will be able to request in IPv6 address. Right now the registries are working with IANA on the decision. There's a proposal on the table that will soon go to informational RFC. ISPs will be able to request IPv6 sub top-level aggregators.

COOK Report: Where can people go to look at these ongoing discussions?

Bound: A good place to start is the Compaq IPv6 Web pages at http://www.digital.com/info/IPv6/. The Next Generation transition working group within the IETF is responsible for designing the mechanisms and procedures to support the transition to the Internet from IPv4 to IPv6. Additional information, including instructions on how to subscribe to the ngtrans-mailing list, can be found on the ngtrans-home page. Pointers to other IETF drafts and IPv6 related URLs will be found in the text box accompanying this article.

In 1998 we have begun the processes by which people will be able to request IPv6 addresses. I will predict that in the second half of next year you will see all the Unix vendors actually ship IPv6 products.

COOK Report: If I am a network manager inside an ISP or a corporation, what kind of coordination issues do I have to think of in planning a conversion to IPv6?

Bound: It's a lot easier than you may think. In IPv6 every address has two associated variables. One is called the valid lifetime and one is called the preferred lifetime. So the valid lifetime of a given IPv6 addresses may be good for ten days. While the preferred lifetime of the same address is set it at only five days. At the end of that five days a timer must go off which says the address is deprecated and you should now begin to look for another address to replace this one.

COOK Report: But how then would this impact on the global routability of addresses?

Bound: Well there is top-level aggregator in an IPv6 address and that top-level aggregator is globally routable.

COOK Report: Right. You mentioned that. But if I'm a corporation with IPv4 addresses taken from my ISP's globally routable block, and I want to change ISPs, I may have trouble getting those addresses globally routed. What will be the analog to this in the IPv6 world?

Bound: In such a case you have a staircase. The low order bit's in your address will remain the same and routability of those bits will stay the same. The prefix will change (high order bits). So if I have an address and my preferred timer goes, I start looking for a new address. Low and behold I get a new prefix from my new provider. So I let the old address time out and then start using the new. And guess what, the new prefix will tell the routers to send your packet to Sprint and not MCI if what you of done is change from MCI to Sprint and are now using the Sprint prefix. By the time the new connection comes live to my desktop, my whole routing infrastructure has already made the change.

Now there is no replacement for conscientious systems administration of the addresses. Someone does have to set up address lifetimes from the router advertisements or from the DHCPv6 server. Someone has to set up the expected prefix at the entry point into the autonomous system. The global routability is clearly defined from the end system's point of view as the entire prefix and, as you go up the levels of hierarchy within the autonomous system, everything looks at the high order prefix of the new service provider.

So in the answer to your question about just one department or division of company changing to IPv6, you could rearrange (by giving it a new prefix) the part of your autonomous system that gave service to that part of your compa-

ny wanting to change. It would be transparent to the end user. The ISP would advertise through its router a new prefix. The company's autonomous system would pick up that prefix and then Cascade that now. The question becomes how fast can that be done? We don't know. I think it's a matter of days however -- weeks at the most. As such you would be very different from the six months to year that would be required for comparable renumbering with IPv4.

COOK Report: Let's try to generalize what you seem to be saying. Could you consider these very large addresses to be rather like freight trains with interchangeable engines? There are various kinds of engines (aggregators) that you put on them to shunt them around at various levels of a hierarchy. Add the if my metaphor holds, what you are telling me is that the different kinds of engines that you can attach and detach very easily to what is still a very very long freight train. But moving such a "train" around in these areas within these hierarchical levels is much easier to control and can be made more automatic than anything that you can do with the shorter IPv4 addresses.

Bound: Your image is indeed accurate and very well phrased. We have had some very bright people working at IPv6 and we are at last really ready to go with it. The big question is what will it take to get large XYZ bank, or semiconductor company or manufacturing company to make the change? Those who adopt IPv6 will face a transition period. But we've built mechanisms for making the transition into the just about any kind of architecture, and we are ready to explain to those companies that wish to give making this transition serious consideration.

With IPv6 you will get the security of IPsec. You will also get quality of service capabilities that you cannot now get in any other way. Mobile applications are also much improved. Now the point is many potential IPv6 users haven't moved yet to the technologies that IPv6 will enhance. So what may make IPv6 deploy, in addition to running out of address space, is that they realize they need to start using these new technologies and they see the problems with NAT. In short there are a multitude of reasons that are converging to make businesses want to change to IPv6 and want to do so immediately. We have businesses right now that fall into this category. But we can't give it to them yet.

COOK Report: Why not?

Bound: Because we cannot yet give them a real IPv6 address. They have to be able to get such an address from their provider.

COOK Report: But if you are going to do a tunneling implementation you wouldn't need and IPv6 address given to you buy your provider. Is that correct? And what is a situation where I would do tunneling versus a situation where tunneling would be of no use?

Bound: yes. I think IPv6 will have to do some tunneling because the entire Internet will not converge on IPv6 simultaneously. The thing is that, if you want the benefits of IPv6 on your intranet, I believe that you will be able to start purchasing from the Unix vendors next year. It appears that all router vendors currently have some implementation of IPv6. I think 3Com is very aggressive. And Cisco right now is one of the premier players on the IPv6 test bed -- routing IPv6 very well. Microsoft has research activity underway and I think at some point we will see that on the desktop.

COOK Report: Well if you do not tunnel, you presumably get more efficient use of your resources?

Bound: Absolutely. Tunneling also creates a management problem. If a packet is tunneled five or six times in a row, your control over it vanishes. But that's true for IPv4 is well. Now clearly you can build a network of IPv6 sites and begin to get your network to those sites. That will happen. Some provider will get a top level aggregator. (We assume this year the people will begin to request them.) When one does, others won't be far behind. We've gotten word from Europe that some PTTs want to start experimenting with IPv6 tunnels between them.

So if an end-user wants the benefits of IPv6, and that end user is willing to do tunneling across the Internet the other IPv6 sites, then I would predict that they should actually start getting the vendor kits out there today. By the end of this year there should be actual IPv6 applications that are no longer considered beta.

The Transition Process

I think we have addressed all the real issues. The routing protocols, which were holding us up, until about four months ago, have now been addressed. We have BGP4 for which supports IPv6. Gate "D" will provide public domain implementations of IPv6 in the near future. They are then going to need some transitioning. The IETF transition working group is providing some very good specifications for this. For example Brian Carpenter and Cindy Joung have a proposal out there that will show how to take the inside of an IPv4 multicast packet and have it multicast IPv6 neighbor discovery packets. I think that, over the next two IETF's, people will build all the necessary transition mechanisms and vendors will agree to implement them.

COOK Report: And the definition of transition mechanism is what exactly?

Bound: There is an IETF working group called ngtrans which is headed up by Bob Fink of LBL, and Tony Hain of Microsoft. There are basic parts of the architecture that inherently support the transition. The architecture is also designed for ease of coexistence with IPv4 which is going be around for long time. We have designed a set of mechanisms that can be used to assist in the transition process.

COOK Report: Are we hearing you say that the transition working group is mapping out a set of procedures or guidebook's that can be used in making various kinds of transition's from IPv4 to IPv6?

Bound: We are actually going to have a set of proposed standards to map out the mechanisms. As far as guidelines of procedures are concerned, I doubt that you will see them from the IETF because, as I said earlier on, what we've learned from previous transitions is that you cannot define a set of procedures that is appropriate any particular transition. Vendors with specific IPv6 products will very likely provide a set of guidelines and transition procedures for use with the product itself.

COOK Report: So the task seems to be one of mapping out how to get from here to their according to what hardware your using, what operating systems, what industry you are in and what your application priorities are. Is that an accurate assessment?

Bound: Yes it is. There will be various scenarios.

COOK Report: Where do things like multicast come in? I'm told that in the IPv4 world one of the problems multicast is that they do not know how to a charge for it when it crosses a network boundary. Does IPv6 any solutions for something like this?

Bound: I think we've solved the scalability problems of multicast and that now is mainly a matter of getting it implemented and deployed for v4 or v6. Billing and charge back are problems that really do need to be solved. But not clear that they're solvable by a protocol.

COOK Report: Do the reasons for using multicast in an IPv4 world change significantly in an IPv6 world?

Bound: No. In fact they are expanded. IPv6 does not use broadcast technology. It uses unicast for the point-to-point transmission of things like private mail and multicast to go from one sender to multiple addresses but not to the entire internet. IPv4 has to use the time-to-live value to scope how far the address goes. It is not able to base action on the address itself. The prefixes on the other hand that are used for multicast in IPv6 are scoped so that when a message goes out, it goes out only where the prefixes tell it to go. The destinations could range from someone else on the user's site to the other side of the Internet.

COOK Report: Could you have some graphical user interface by which it the sender of the message could easily determine exactly where he wanted to go within his organization or outside?

Bound: You could indeed do this and it is something that would be impossible to do with IPv4.

For IPv6 related URLs, see p. 11 above

BBN's Blumenthal Describes 1995 - 96 Regional Network Integration Process
Explains Details of Current OC-192 Sonet Buildout

Editor's Note: With the amount of consolidation going on in the Internet and especially within WorldCom, one's ability to knit together, seamlessly, the technical operations from different acquisitions becomes a matter of higher and higher priority. Steve Blumenthal has been with BBN, now GTE, for more than 21 years. For most of that time he has been working on the Internet. With the problems of melding acquisitions together in mind, we asked him to elaborate on the general nature of his experience at GTE and lessons learned from the amalgamating of BARRNet, SURANet and NEARNet.

Blumenthal: I have been involved with the development of the Internet for most of the time that I have been at BBN and now GTE, both in R&D and network engineering, deployment and operations. I was involved with the development and operations of the Atlantic SATNET, the DARPA Wideband Network, and the Defense Simulation Internet. In 1989, I was part of the group that launched NEARnet, with my group having responsibility for operations and maintenance. In 1995 our group won a contract from America Online to build out and run a significant portion of their nationwide dial-up network. In 1996, when BBN decided to focus all of our energy on the Internet market, I took on responsibility for network engineering for all of BBN, including BBN Planet. At this point, BBN had acquired NEARnet, BARRnet, and SURAnet and was in the midst of consolidating these regional networks to create a national network.

Integrating Acquired Networks

COOK Report: In the context of Verio, WorldCom, Mindspring and Earth Link, all of which are doing rather many acquisitions, it would seem that your experience in building a national service from the integration of these three networks would be highly relevant. What were some challenges you faced?

Blumenthal: There were quite a few. Each one of these networks was designed and run by very good people. But each network differed from its counterpart in significant ways. Also, the Internet was itself in the midst of a major transition from a period where most of its users were universities to a phase where most of the users of the Internet were businesses.

COOK Report: This was in 1994 and 1995?

Blumenthal: Yes. While all of these networks were designed by very smart people, they were really designed to meet the reliability expectations of the university community.

COOK Report: And businesses would have different reliability expectations?

Blumenthal: Right. Although each of them, NEARNet and BARRNet in particular, had quite a few business customers. The business customers tended to represent the research side of their parent companies, such as HP Labs and Xerox PARC. They were not yet the kind of user that a company like Cisco or Dell Computer would become. They were not yet booking or three billion dollars worth of orders over the Internet with their prosperity critically dependent on their Internet connection. When users like Cisco, Dell, large commercial banks, and other businesses began to appear on the horizon and use the Internet for commerce, it meant that we had to change the design, the build out, and the support infrastructure for these networks.

When we took over NEARnet, BARRnet and SURAnet, each was an independent Autonomous System, and each one had its own independent network engineering and operations staff. One of the first things that we did was to consolidate all the network operations up here in Cambridge. A lot of the SURANet and BARRNet operations people stayed on. But when they did so, they assumed different roles in our organization. A number of them moved into network provisioning roles, making them responsible for bringing new customers online in their respective areas of the country. While some did move to Cambridge, most stayed where they were and changed jobs. Because people with these networks had grown up with the Internet, there was a lot of Internet knowledge in those places. Quite a few of the operations people into engineering. With the Internet linking us all together, we found that the engineering work could be done from any part of the country.

COOK Report: Well in College Park, MD with SURANet and in San Francisco area with BARRNet, there would have been network operations centers. How which you describe the changes that affected these centers?

Blumenthal: These changes happened in the 1996 time frame, about a year after the acquisitions. Network monitoring and customer service were moved to Cambridge. Most everything else stayed in place -- including regional engineering groups, regional field service, and regional customer provisioning operations. In Maryland we still have a large customer provisioning group and that group works in tandem with our customer provisioning group in Cambridge. We have field service centers in Maryland, Palo Alto, and in Cambridge as well as in other parts of the country.

COOK Report: You are presumably going to have a bunch of routers in those locations.

Blumenthal: Right. And in the engineering group we have a lot of people who have advanced in their careers by moving from operations into engineering. Engineering is responsible for network construction and these construction activities are pretty distributed. We have people in Palo Alto, Maryland, and Cambridge who are involved in infrastructure provisioning -- doing the project management work in building new pops, in other words.

COOK Report: Well what would be some of the other areas of divergence for your acquired networks? For example, there might be different accounting systems and different customer care systems. What would be some the areas in which you would have to take independent functions, and check to see that they are running in similar ways?

Blumenthal: Most of the standard functions: billing, customer care, network monitoring are all run centrally out of Cambridge. Today we use some other regional offices to do some network engineering, debugging of problems, and keeping the network alive. We have divided the network up into regions and assigned various people regions of the network. For example if we are doing a big upgrade in Atlanta, that's probably going to be done by the network upgrade group in Maryland. If we're doing a big upgrade for the western part of the country, that's probably going to be done out of the office in Palo Alto.

COOK Report: What about things like customer billing systems and various billing costs in procedures that would exist independently with each of these three networks and yet would need to be blended into a single uniform system over time?

Blumenthal: By early 1996, we had, moved substantially all of those things into a single customer care and billing system. We still had the different contracts that customers had signed with our acquired networks, and we had some variability there.

Standardization

The big activity that we undertook in 1996 from the engineering side was the development of a standardized configuration for our POPs. We decided to close some POPs in some small markets, but we created new standard designs for the ones that we wanted to preserve and upgrade. This included a design for backbone POPs and another standard design for regional or metropolitan area POPs. NEARnet, BARRNet and SURANet, were initially linked together using the COREN backbone to provide interregional connectivity.

COOK Report: COREN was a very shrewd deal that you put together with MCI, correct?

Blumenthal: Yes it was a good deal that had been worked out by the NSFnet regional networks and MCI. In 1996, we built out a nationwide DS3 backbone and built new POPs in LA, NY and the midsection of the country. We built new pops in places like Houston, Detroit, Cleveland, Chicago, Los Angeles, and Dallas -- places that were outside the original NearNet, SURANet, and BARRNet geographical regions. While we filled in the middle of the country, at the same time, we went into our three regions and standardized the design of our existing pops so they all had the same configuration of router equipment, and DSU/CSUs.

We then standardized all the routing. Prior to 1997, we were still running separate Autonomous Systems in these three regions. During 1997 we moved everything into AS-1. In late 1997 we acquired Genuity, a major web hosting service company. We are now in the process of moving their Internet traffic onto AS-1.

COOK Report: What are some of the difficulties of running multiple autonomous systems that could motivate you to undertake this amalgamation?

Blumenthal: When our engineers needed to debug a routing situation a particular segment of the network, we found it highly desirable that they have a common understanding of the nature of the router configurations with, as well as the relationships between the customer's routers and our routers. A major area standardization was BGP and the policies surrounding BGP routing. We put a lot of effort into reducing the operational load of BGP.

COOK Report: Would you give an example?

Blumenthal: We have a large customer on the West Coast that had been a customer of BARRNet before we purchased them. Now this commercial customer in California has a main data center in that state and they created a backup data center elsewhere in the U.S. They wanted to be able to switch from one data center to the other seamlessly should their network service ever be interrupted.

We needed our BGP routing to be done in a uniform way throughout our network because we needed to be able to debug a problem in one part of our network from any other part of our network. Without a uniform approach to BGP across our entire network, this would be impossible. For example, if we have a problem in a router in Atlanta and our East Coast engineers work on it all day long, when they begin to get burned out, our West Coast engineers can jump in and finish the task.

Let's go back now to the case of our West Coast customer with two data centers. This customer wanted the data centers to appear to the Internet as identical and interchangeable. The customer insisted that any interruption in Internet service not be noticeable to its customers who were accessing data in these data centers. This customer also had some very interesting security features impacting its access to the Internet. Now they had an internal link between their data centers that would make switching operational centers very easy. In the beginning of 1996, our BGP routing policies were not uniform across our network and there were differences between the way that we interfaced with their California data center and their backup data center. Therefore, when we had to do a switch-over from one data center to the other, it was not seamless because it required manual changes to routing tables. Our customer did not like this.

This customer's displeasure, which we moved to rectify as quickly as possible, became a good example of the problems that could flow from non-standardization if you are trying to sell a nationwide or international service. We have a number of customers which have bought multiple connections from us for the purposes of backup or even just for load sharing. Digital Equipment (now Compaq), for example, buys connections from us all over the world. They are doing a lot of internal order-entry processing from sites scattered around the globe. They want to run standard business processes at all of their locations, and they don't want to have to make changes to their processes because our internet service doesn't provide them uniform consistent access at all of these locations.

Multi-tasking as a Precondition for Growth

COOK Report: Can you give any kind of indication of what allocation of resources it was necessary for a company such as yours to devote to the development of new uniform systems, while, at the same time, having to keep nonstandard systems operational? Also how did you migrate customers between old and new systems?

Blumenthal: It happened in two phases. For the customer facing stuff we would migrate groups of customers according to the topology of their connection to our network. Customers all connected to a given POP would be migrated at the same time. We would also try to migrate groups of POPs according to the region of the country in which they were found.

One of the things that we were doing in this period was to come up with a consistent way of doing billing. Some these networks, I think BARRNet and SURANet, had fixed-price billing for a T1 connection. NEARNet had pioneered usage-based billing where we would measure how much you used the Internet. Each month we would send the customer statistics on their usage of the Internet. We now have a system called Stats Advantage where the customer can log into a web site and check their own statistics. In fact, we are now moving to web based customer service systems in general for things such as checking trouble tickets. We have rolled this out nationwide on a region-by-region basis.

COOK Report: Well, during this period of time you are doing a lot of duplicative things in parallel. What percentage of your overall effort did this kind of multitasking cost you? While you have staff run one set of systems you have to have other staff developing and coordinating new systems and moving current customers onto them.

Blumenthal: Your question is a relevant one and highlights a real problem. In 1996 and early 1997 the standardization efforts had a major impact on the business we did. For a while some people found themselves with more than one job. In the first six months of 1997 we focused a very large proportion of our engineering resources on the solution to this problem. While this investment of these resources slowed down our growth during that period, we feel that it has more than paid off in that we have now become an industry leader in quality of service. We have had an independent organization measuring our customer satisfaction on a monthly basis. We noticed a very dramatic improvement in customer

satisfaction after we completed this project in mid-97. In an industry where there is a tremendous amount of churn we have had very high customer retention.

COOK Report: So how do you react to the following conclusion? As one is faced with the problem of coordinating network acquisitions, one has a strategic choice to make. Either you must have 20 to 30 percent or more of your staff that has to be devoted almost entirely to developing the uniform systems and transitioning customers to them while the rest of your staff keeps things running on the prior basis. If you don't have the luxury of allocating staff in this manner, you will find that your only alternative to letting your service fall apart is to dramatically slowdown your new customer acquisition and growth.

Automation of Systems and Standardization of Routers

Blumenthal: Those are reasonable conclusions. Our investment in these changes created a platform for scalability. Prior to this time, many of the activities that were done in BARRNet, SURAnet, and NearNet took place in a very manual, labor-intensive way. One thing that we did during this period was to create a large number of automated systems. For example the maintenance of the database indicating what customers are assigned to what ports on our routers. This used to be a manual flat file that was updated by an army of clerks. Having such a file increased to the time it took to provision customers. Moving to an automated database for this was one of several things that we did which resulted in improvements in the time it took to provision new customers, the accuracy of the data, and reduced the labor necessary to do so.

We went through it and made sure that we had mechanisms in place so that when data was put into our central database, there were all kinds of cross checks and consistency checks so that people couldn't inadvertently put bogus information into the database. The data was entered once and then this database fed information to our operations and customer service systems. Having the database much more accurate and standardized helped the operations staff be more responsive. When a customer called in with a problem they would immediately be able to look up the correct circuit ID, router port configuration, etc. for that customer. They would also get a list of other services that a particular customer had purchased from us. For example, in addition to Internet access, we might also be managing their firewalls and hosting their web servers. We put considerable effort into developing these back end systems that helped our operations staff to be very knowledgeable about each customer.

COOK Report: When you look at network hardware like routers, do you find it advantageous to standardize on one kind of hardware throughout your network?

Blumenthal: Yes. For our backbone routers, we standardized on the Cisco 7500 series in 1996. Fortunately, BARRNet, NearNet, and SURANet had all picked Cisco as their router vendor. However, they were all were running slightly different versions of Cisco software and they had different types of Cisco routers. In the case of BARRNet, I noted that when I first visited one of their POPs, they had some 7500 series routers, but they also had a number of 7000 series routers which was an older version. They even had a few Cisco AGS+ routers, an even older version. We systematically went through and upgraded all the older equipment and deployed the new equipment in a standard configuration.

Another thing that we standardized was the local network interconnection between the routers inside the POP. At that time we came up with a highly fault tolerant design. We have stuck with this basic fault tolerant design, but have upgraded the inter-router interconnection speeds.

Let me also give you one more example of automated system. When we came out with our usage based billing structure, we needed to collect accurately statistics from all the routers including customer premises routers. What we found surprised us. Some of the routers did the calculations accurately and some didn't. For the most part these problems were cured by software fixes. BBN historically has had a very strong network analysis capability. The guys who do this for us really dug inside the Ciscos to find out how they made their calculations. We made a lot of recommendations to Cisco as to how they could improve their router's capabilities in this area.

Another area that we found we had to standardize on was out-of-band access. If, for some reason, we could not use the Internet to reach our routers, we had to have a separate means of dialing directly into them. We found that we had to standardize this as well because we didn't want to have different flavors of dial-up modems with different password configurations. This sort of thing just drove the operations staff crazy.

COOK Report: It certainly sounds like that as long as major growth is continuing in the Internet any major company would have to have between 10 and 20 percent of its staff that is continuously focused on these issues of long-range planning, coordination, and troubleshooting. Is that a reasonable conclusion?

Blumenthal: I think it is reasonable. I don't know the exact percentage devoted to those areas is, but I do believe that it is higher that the resources allocated by our competitors.

COOK Report: So in thinking about the business of being a national service provider, one needs to remember that necessity of having a planning provisioning staff as well as folk who provide your day-to-day services?

Blumenthal: Absolutely. There is another place where standardization across our network helps and that is in our peering relationships. We have both public and private peerings with other major providers. We need to make sure that our interaction with those providers in Palo Alto are the same as they would be in Dallas or Cambridge. Standardizing our routing by moving everything into AS1 helped this process. When we were running separate autonomous systems, it was difficult to balance the sharing of traffic. While, right now, we are all using shortest exit peering, I suspect that over time we will evolve to a more sophisticated form of best exit type routing between the peering partners. We have increased the the number and size of our private peerings interconnects to the point where we now have several OC-3 interconnects.

COOK Report: From what you been saying is it safe to assume that operating two different backbones each with a different kind of router is not an efficient use of resources?

Blumenthal: Correct. Using to peak efficiency the equipment of each different manufacturer involves a major investment of engineering time. This issue is also relevant to the ATM world. Because if you have an ATM infrastructure, you have to have a set of people devoted as specialists to this technology and to understanding the ways in which that it can engage in complex and unexpected interactions with your routing.

Maturing of the Internet

As to the question of having different routers, while I would love to have a second source, because I am very concerned about being too dependent on Cisco, let me say that we had worked very closely with Cisco to redesign the service that they give to Internet Service Providers. This is another one of the maturing of the Internet phenomena. In the early days of the Internet, when it was very focused on the academic community, there was a very very tight collaboration between the leading technical people at the major ISP's and Cisco. When a routing problem developed, they would send Tony Li an e-mail message and get back

patch from him most likely later that day. Now when Tony left Cisco, and as the Internet has grown and matured, and Cisco as well has matured as a company, that intimacy broke down.

This was one of the problems that we were experiencing toward the end of 1996. At that time we worked with Cisco to help them redesign their service strategy for Internet providers. The outcome has been a dramatic improvement in the level and quality of service that they are providing to us today versus the state of things in late 96. It took them six months to work out all the kinks but we're now back to the point where we can get prompt and adequate support when something goes wrong.

In the old days when we were dealing with problems at the Tony Li level inside of Cisco, there was really no independent means of tracking these problem reports within Cisco. With Tony there it wasn't necessary. Problems were getting fixed quickly and we were able to keep things running. The process of the Internet engineer talking directly to the Cisco developer was going on independently of any action on the part of the management of this company. But we have now been able to institutionalize this kind of personal support within Cisco. We now have people inside of Cisco who know our network inside and out and work with us on a daily basis. And yet Cisco has also created a system where, if we have a major problem that is not immediately resolved, John Chambers will know about it rather quickly. He finds out about it not so much by means of a call from myself or George Conrades, which was the case back in the end of 96 early 97, but rather because his people within Cisco are keeping him informed.

The Current Build Out

COOK Report: Can we move onward now and talk about your major network build out that's currently underway?

Blumenthal: Yes, first let's talk about our fiber infrastructure. We are in the process of moving from a situation where we lease circuits and lease colocation space from several carriers to a new role as a facilities based Internet services provider. We are building out our own fiber infrastructure and our own pops. We're beginning to move equipment into those pops, to light up the fiber and to use circuits derived from that fiber.

COOK Report: Qwest is putting the fiber in the ground, is it not?

Blumenthal: Yes. We have purchased 24 strands of dark fiber Qwest. As Qwest puts the fiber in the ground, we are going along behind them and lighting it up by attaching the opto-electronics to it. We are installing the dense wave division multiplexing and SONET equipment in new pops located near their railroad right-of-way. We have also purchased fiber from other providers and have lit up some of that fiber as well. When this construction is done, we will have about 17,000 route miles of fiber. Some of this is fiber that we inherited from GTE. Some we have purchased from other providers. The majority is purchased from Qwest.

We have lit up capacity now from San Diego all the way up the West coast to San Francisco and across the country to New York City. And that we also have lit our fiber from Boston all of the way down to Washington D.C. We have also lit from Kansas City down through Texas all the way to Houston. We have Internet circuits running on the West and the East coasts and are in the process of bringing them up cross country. We will run OC-3s and OC 12 on this new fiber this summer. We are in the process of moving our IP equipment into a number of major new pops we're building this year. The first will go live this summer in Los Angeles.

COOK Report: When you talk about the 24 strands of fiber, that had an announced price of 500 million dollars did it not?

Blumenthal: Right, and we added some more to that -- namely a whole route in the Southeast into Florida from Qwest.

COOK Report: Well you have a lot invested in the fiber itself, but in addition to that, you also have a whole lot invested in the opto-ectronics. If you're talking about 500 million for the dark fiber, presumably that does not include any money for the opto-electronics? One has to wonder about the total amount.

Blumenthal: It is very very large. About another 500 million dollars or so.

COOK Report: That's impressive! Someone else said in that GTE earnings had been taking something of a hit because of the size of the investment that you are making now and will have to continue to take such hits for several years into the future. But given current technical realities, it would seem that you have little choice but to do what you are doing. In short what do you have to do then to deliver the bits under the new fiber regime?

Blumenthal: There are a whole array of services that BBN was offering or GTE had in the planning stages prior to the acquisition. Since then a lot of this it has been brought together within the GTE Internetworking. We have the responsibility for the build out of the fiber and really being the wide area network part of GTE.

We start with the purchase of the fiber. Then on top of that we have selected Nortel OC-192 SONET opto-electronics to light up the fiber. Nortel has announced that this equipment will support 16 dense wave division multiplexing channels on each fiber. We are also deploying SONET ring protection so that we will be able to cut over to another fiber in a few milliseconds if we have a fiber break. On top of this we're going to offer commercial ATM and frame relay services to business users. This is basically to replace some other carrier services that GTE is currently buying. We will also use it to support our America Online contract which is currently running on someone else's frame relay. We have just received an extension of that contract both in terms of time and volume.

The Fate of SONET

COOK Report: Since you mentioned SONET, let's verify the issue of SONET's near term fate. Although you can do TCP over light without SONET, the ironic thing about dense wave division multiplexing is that it so extraordinarily increases the carrying capacity of a single glass fiber that for the next few years, until a single organization's IP traffic can fill several multiples of OC 192, that organization will want to sell slices of available bandwidth from that fiber to other carriers. The only way that it the owner of the fiber can do this it is by means of the purchase and application of SONET multiplexing and de-multiplexing equipment. Even though you don't need it to run an IP network, you still have to buy it in order to be able to sell to someone else the unused capacity of your fiber. Is that correct?

Blumenthal: You are correct. Certainly for traditional circuit services we would keep the SONET equipment. Now at the recent InterOP Cisco demonstrated OC-48 SONET interfaces on their GSR routers. You may certainly assume that they are working on OC 192.

COOK Report: As IP traffic grows, you will eventually fill an OC 192 and then several multiples of an OC 192. When you have finally filled a entire glass thread with IP traffic, you will do longer need SONET equipment on that thread in order to sell a fraction of the bandwidth to someone else. Is this a correct conclusion?

Blumenthal: Yes. There might be some pieces of glass that we would devote exclusively to IP router-to-router traffic. In that case you would not need SONET equipment on them. I think that you may see some of the protection features provid-

ed by SONET incorporated directly into the routers. It really will make the need for the expensive SONET equipment go away. However as long as you are provisioning T1, T3, OC-3 and OC-12 circuits from this fiber, you would probably need the multiplexing capability of SONET.

COOK Report: And right now there is enough demand for those circuits so that if you are able to sell a large number of them, you can use the resulting income to recoup part of your original 500 million dollar investment in the dark fiber. Therefore, paying for the SONET equipment that makes this possible to do is no big deal?

Blumenthal: Right.

COOK Report: But the continued necessity for investing in SONET is one reason why a drastic collapse in the cost of circuit prices is unlikely?

Blumenthal: That is probably true. Given the difference in the growth rates of voice versus data, there is no doubt that data is our future.

COOK Report: Well let's look at where this seems to be going. If you are running an IP network over glass, you are running a connectionless network. Now if that glass gets cut, having the self-healing capability of a SONET ring is all very nice. But in theory, at least, in a connectionless network, should you not be able to recover from a fiber cut by simply doing rerouting?

Blumenthal: That is correct. In such a case you might not need SONET. It becomes an extra overhead the layer -- rather like wearing a belt and suspenders.

Build Out - Further Details

But let's go back to be specific factors of our build out. As I said above the Nortel equipment is in the process of the deployed. We are lighting segments of our network and building pops. We are installing in those pops the new Cisco GSR routers. We have worked with Cisco to develop a new version of the GSR, the 12008 which is a lower power and smaller package that fits in the with the power and heat emission constraints that co-location in telephone pops requires. The original version of the GSR was simply too big and too power hungry. We are also beginning to place Ascend modems in our new pops as well as are older ones. We are installing a tremendous number of modems every month both for America Online and for our own dial-up network, part of which is used for business customers and part of which is used for GTE.net, a major consumer Internet service offered by GTE. All of this traffic will be moving onto our new network over the next two years.

We are also putting in place the ability to connect up to ADSL service which GTE is rolling out aggressively in their central offices and in some out-of-franchise areas as well. We have deployed and launched IP fax services and we also have some other voice over IP services coming. We will be putting that equipment into our pops and we expect to have about 30 cities up running for IP Fax service by the end of the year. We have picked Ascend-Cascade as our supplier for ATM and frame relay service equipment. We will be providing interlata nationwide ATM and frame service later this year.

COOK Report: What choices will you face what you want to move beyond 16 lambda's in your wave division multiplexing? How do you think about planning for such an eventuality? What is involved? A total replacement of equipment?

Blumenthal: I don't have the figures at my fingertips but I believe that something less than a complete replacement of equipment is involved. Very likely the amplifiers at the end of the circuits would need to be upgraded. I believe the upgrade involves software and hardware including some components in the terminals would have to be upgraded. The amplifiers along the path should have enough bandwidth to be able to handle more wavelengths, because, as they go to more wavelengths, they are packing them closer together. Therefore it is really the detectors and transmitters that need to be more narrow.

International Activity

COOK Report: What you doing internationally at this point?

Blumenthal: We have a partnership with Equant which is the commercial arm of SITA. SITA is the international consortium of airlines that operate a worldwide frame relay network. We are using their worldwide network to provide connectivity outside of United States. In doing this we are primarily focused on selling Internet services to U.S. based multinationals.

COOK Report: if Equant wanted some across U.S. bandwidth, it seems that you could very likely provide it in return for bandwidth outside the US?

Blumenthal: Perhaps. They have not asked us. They recently bought Sabre network from American Airlines. They basically give us international infrastructure without having to build it. In return we provided the IP services that they did not have and which they could then go out and sell. Their network goes into any place in the world into which you can commercially fly. They have copper. They have fiber. They have pops in 220 different countries. And they have a lot of people on the ground who can go in and fix things when they go wrong.

We are also building major pops oversees. London, Amsterdam and Sydney are up and running. We have several additional POPs in Europe, the Asia-Pacific region, and Latin America under construction. We use these pops to collect local traffic in the various parts of the world, and then by means of local peering hand this traffic off to other Internet providers in those regions.

We are also making significant invests in undersea fiber cable capacity. We have made two purchases. First, a substantial amount of bandwidth in AC-1 in the Atlantic. We have also purchased capacity on America's II cable into Latin America via the Caribbean. We are also in the process of making additional purchases in transatlantic and transpacific capacity.

COOK Report: Any comment on Project Oxygen at this time?

Blumenthal: We are looking at that and we may make some additional purchases

Contemplated Names Council Putsch: Is This Postel and ISOC'S Understanding of Democracy? Does IANA Do Open Processes?

[**Editor's Note**: These comments were based on a Position Paper released on August 8 by a group calling itself the World Wide Alliance of TLD Registries. What follows is a revised and expanded version of an analysis that we published on the net on August 10, 1998.]

Jon Postel, apparently speaking on behalf of the United States, has endorsed a proposal that would create an IANA Names Council made up solely of the country code TLD administrators whom HE appointed to the positions they now hold. In return this Names Council states that it endorses Jon's draft by laws for the operation of the new IANA corporation. It adds that it will speak on behalf of ALL DNS registry efforts. NSI, OPEN RSC are not invited to participate. We are forced to conclude that unless NSI applies to join, Dot com registrants are to have no representation.

Later in the body of this message we examine the interesting collection of people, many of whom are Jon's friends, who propose to decide

policy for the rest of the world. Let the record show that Jon endorses the proposal of Howard, Chon and Turcott as the administrator for the .us country code top level domain. Readers may decide for themselves whether they are pleased by the way in which Jon represents the interests of his native country.

This all seems a bit bizarre. Therefore we repeat: this self-appointed group of 90 of the 220 country code administrators is gathering together pledging their fealty to Jon and saying that they are now the Names Council and therefore that they will also make policy for the seven generic top level domains (.com, .net, .org, .edu, .mil, .gov, and .int) as well as for their own country code domains. Presumably if NSI wants to join, it may do so by pledging loyalty to Jon and to ISOC. Then the 2.5 million .com registrants will have a single vote the council -- a vote equal to that of the administrator for the Pitcairn Islands who lives in Switzerland.

This subset of 90 country code administrators (all 220 have one million registrants compared to 4 million registrants in the seven generic domains) is suggesting that, with a total of 20 percent of world's registrants, it will make policy for the entire world. Since Postel is a signatory on behalf of .us, we must assume that this outcome represents his solution to the problem of the commercial Internet. Ignore the developments of the last decade and run things as he always has. But the old way of doing this, leaving the world's most critical communications infrastructure in the hands of the good old boy network of engineers who built it will no longer work. That is what this war over the new IANA is all about. ISOC apparently will go to any end to preserve its and Postel's hegemony. Consider the events of June 10th.

Overwhelm your Enemy

Then at least Don Heath spoke his mind. On June 10, 1998 in a public message to his ISOC supporters, he slammed the emerging International Forum on the White paper process and suggested that if ISOC couldn't sink it, it would seek to "overwhelm" it. When at Reston, ISOC found out that IFWP was real, it sought the to overwhelm it by joining IFWP. Heath took a swipe at Jim Dixon in early July. Trying, but failing, to cut him out of the picture, ISOC participated in the second IFWP meeting in Geneva. There Fay Howard put forward a RIPE plan to bring country code top level domain registrars together in a names council to represent country code interests.

After the Geneva meeting ended, when a plan for a final wrap up meeting was put forward on the steering committee, Heath came out against it saying that one was not needed. The strategy was obvious. Without such a meeting ISOC could declare to the world that it was acting on behalf of the World wide Internet and implement Jon Postel's by laws for the new IANA corporation. The same by laws drafted and imperiously handed down 'ex-cathedra' by Postel to considerable criticism. Jon last week put out a second version featuring an IANA board responsible largely to no one that was judged as even worse.

Wrap up Meeting and Consensus by-laws

In the meantime at the end of last week two very important things happened. The steering committee of IFWP voted 11 to 8 IN FAVOR of a wrap up meeting….presumably to be used to arrive at by-laws for the new IANA corporation. Those voting against such a meeting were Heath and Educause's Mike Roberts and a European group representing POC and CORE. ISOC seems to fear exposure. For at the request of one or more members of the steering committee the size of the vote and the names of those voting against have been withheld from public scrutiny. I have dual sourced these statements with steering committee members.

At the same time on Thursday the 6th of August NSI posted its effort at a consensus-based version of IANA by laws. This version has been well received by virtually everyone outside of the Postel/ISOC/CORE orbit. This was a brilliant move by NSI. NSI should realize that it now has an opportunity to make second positive move. If it joins Open ORSC, the save for CORE all the global TLD interests would be united. Core would of course be welcome to join. At that point the Names Council would have self-organized.

However as the Singapore meeting is about to begin, is quite troubling to note that the country code group now advertising itself under the grandiose title of "World Wide Alliance of TLD Registries" has been put on the agenda to present its case. We are talking about an allegedly open and transparent process where a group of Postel's hand picked country code domain people claim to be able to decide policy in the absence of any representation from GTLDs.

Jon Postel Doesn't "Do" Open Processes

We are also talking about a situation where Postel has stated that if a government requests a change in the administrator, he will comply. But we must ask why has he not made an effort to directly encourage national governments to do this. One result, as we note below, is that France has five votes in the World Wide Alliance of TLD Registries to the United States' single vote. Given ISOC's steady insistence on playing the spoiler role, the result this self-defined exclusivity on the part of the so called World Wide Alliance could be a split meeting and the imposition of a solution by the US government.

Readers are invited to decide whether the group that claims below to have the right to become the names council for the Internet world wide is worthy of support. Most who have leapt on this bandwagon have financial interests at stake. Or are anti US and anti NSI in their positions. We call on Jon Postel to acknowledge the embarrassing position in which he has placed himself by removing himself as a signatory from this ridiculous self-serving document. His signature on behalf of the .US domain is particularly inappropriate when there is an ongoing NTIA proceeding on the control and use of the .US domain.

We also point out that Postel is actually a SUBCONTRACTOR for the .us domain and NSI is technically the contractor as part of the cooperative agreement. That calls his recent actions into question to an even greater degree. NSF is the awardee and therefore must be consulted by Postel before he signs such a document as this on behalf of the United States. Perhaps Jon would be willing to show us NSI's or NSF's approval for his signature to this document? We'll wager that Jon has no such approval. His point of view seems to be the Internet, cest moi. Unfortunately he will now have to learn that he doesn't single handedly speak for and set policy for the Intertnet world wide.

During the third week in august others reached a similar conclusion. Michael Dillon wrote on the ORSC mail list: "I would like to see a new IANA become the central point of authority for Internet matters, slowly and uneventfully. In order to achieve this we need more work on consensus, compromise and cooperation. This means that nobody throws their weight around, nobody uses force against the other guys and nobody runs the show. It means that negotiations take place to slowly draw everyone into a central position of agreement." To which Jim Fleming commented: "The old IANA people and the ITAG people seem to have a different approach and have apparently told people this at the IFWP meetings. You will note, they are not part of these discussions. How can you expect any of the above to happen if people do not participate?"

The conclusion of Roeland Meyer on August 22 was even more striking: "Actually, you make a real strong point. If the IANA folk have a discussion list, it is certainly closed to me, as well as, most that I know. It could be argued, by Postel and Co, that they don't have time for such silliness. However, he's not available to anyone else I know either. If he's doing something, and I'm sure he is, why do we all find out after the fact? That kind of blows his credibility vis-a-vis open processes. On reviewing the latest Draft-Postel, this is kind of under-lined (during my feed-back to the NSI review, which Dan and I both did). Jon Postel doesn't "do" open processes.

Let the Record Speak

What follows is a collection of material from the alliance's web page. http://www.apng.org/wwtld/positionpaper.html My comments are interspersed with the material.

Subject: TLD Position paper for your approval Date: Tue, 04 Aug 1998 12:22:45 +0200 From: Fay Howard
To: wwtld@ripe.net

Dear ccTLD Administrators, Following the announcement last week, a draft Position Paper is set out below for your comment and approval. The paper has been kept brief but sets out some important basic points. These can of course be built on at a later stage but we need to start somewhere. We would like to be in a position to announce our unity at the IFWP -Singapore in respect of these aspects of forming a new 'IANA'.

If you agree with the principles of the position paper we would ask that you reply and give you support even if you are not able to attend the meeting of world wide TLDs in Singapore next week. You will also be able to find this document on the Web site of the World wide alliance of TLDs later today . For those of you who did not receive the initial announcement, this can also be found on the Website. Signed, Fay Howard (CENTR) Prof.Kilnam Chon (Korea) Bernard Turcotte (Canada)

COOK: Professor Kilnam Chon BS (Osaka) MS (UCLA) PhD (UCLA) is the Chair, BoF-APAN Co-Chair of Asia-Pacific Coordinating Committee on Intercontinental Research Networking. ISOC's man in Korea. The players are all ISOC or MOU related. Toru Takahashi is secretary general of the Internet Association of japan and one of the most powerful in the Japanese Internet. Close associate of Jun Murai. Strong MOU supporter. Fay Howard (CENTR) is (or was) an officer in Terena -- one of the three founding bodies of ISOC. She was assigned by RIPE to head up the development of a European ccTLD registry. Bernard Turcotte was nominated to serve on the POC last October. Heads ISOC Canada. He is Assistant professor of Microbiology at McGill University. Hardly a neutral group. But lets remember Don Heath's pledge to overwhelm the process

From the Web Site Announcement:

The representatives of TLDs listed below hereinafter referred to as the signatories, [to the

cont'd on p. 24

Executive Summary

Content vs. Carrier Peering War, pp. 1- 11

Three years ago Exodus was one of the second tier national backbones with peering agreements at the public exchanges. Today it is a publicly held web farm company sitting on a war chest of a March 98 $69 million dollar stock issue and July $200 million dollar bond issue. Led by the venerable Ellen Hancock of IBM and Apple fame, Exodus, with only a temporary peering agreement with BBN, has become a national web farm hosting such huge sites as USA Today and GeoCities.

On July 9 Exodus was informed by BBN of a non renewal of its peering. Instead of quietly renegotiating connectivity behind the scenes, Hancock has taken Exodus' dispute with BBN public -- warning her customers in a letter of August 5th that BBN and not Exodus would bear the most pain as a result of the dispute. From what we can determine after an exhaustive examination, it is Exodus that has erred in calling attention to the weakness of its business model. The Exodus model assumes that, since it is now a highly desired content provider, it should be allowed to maintain its no cost peering agreements with global backbones like BBN despite, according to some estimates, an imbalance of traffic with BBN that is something on the order of sixteen to one.

For the past two weeks BBN has been demonized on the NANOG mail list by Exodus supporters and those who insist that Exodus get free connectivity because it has content that BBN customers want. Exodus reacted as though BBN were the only backbone that was breaking peering with it claiming to have transit free interconnectivity with all other providers. We conclude that at best Exodus may have cost free peering with all other majors. But, even if this were true, Exodus has absolutely no guarantee that it will continue past the expiration of its current peering agreements. In reality we suspect a considerable fudge factor. Certainly a reading of Exodus January 1998 S1 on Edgar makes it clear that Exodus was having to pay for some of its connections. The caveats listed by Exodus in its business model in the January 98 S1 are extensive and make for instructive reading in the context of current round of finger pointing.

We publish an argument by John Levine that, based on the assumption that content is now king, justifies Exodus' position. We also publish a discussion with Michael Dillion and Sean Doran where Sean points out the very heavy burden maladjusted TCP web traffic can impose on BBN's backbone performance. Of course if Exodus has the right to transmit its content at no cost, it also will lack the motivation to carefully engineer the behavior of its TCP congestion windows with BBN.

We conclude our article with a 5,000 word interview with John Curran done last Friday August 21. While John was careful to say nothing explicit about any specific BBN peering situation, he offers the most detailed description of both the industry's and BBN's approach to peering that we have seen.

The principles that he explains are quite basic. First that peering in the Internet has been and remains based on the assumption of symmetry in traffic streams. In the case of BBN there must be a ratio of traffic between peers that is not greater than 2 to 1. As a result BBN has upwards of 50 peers. Second, the assumption in the industry has been that both sender and receiver pay for traffic with each paying about half the cost of meeting in what Einar Stefferud recently referred to as the mythical middle. The problem arises that the mythical middle vanishes when peering with a web farm causes asymmetrical traffic. If BBN has to open up half a T3 of inbound bandwidth for every T1 that Exodus has to install, the burden on BBN becomes especially great when hot potato routing means that it goes immediately onto BBN's backbone. If this did not happen and it were Exodus responsibility to carry the traffic flow to an exit point closest the location of BBN's customers, this could be a way of restoring the symmetry destroyed by the web traffic. [The letter from BBN leaked on the NACNet web site made it clear that BBN and Exodus had been experimenting with this approach known variously as best exit or longest exit routing.]

What is only now being recognized is that free peering relationships with carrier backbones and large web farms engender traffic flows that are so asymmetrical that we move essentially to a receiver pays pricing model for the Internet. To quote Curran: "if we enter a world where the senders don't really pay any incremental costs, you face some huge implications. You end up with a situation where, for example, a sender could decide to send you a large video image when you connect up to his web site. . . And I guess I am just a little bit concerned, if not from a business perspective, from a public policy perspective about the multimedia spam possibilities when senders are not paying any real incremental costs for sending more information." Curran suggests that the industry continue to experiment with best exit routing and with settlements based peering where he defines that the content provider would pay only for the portion of its traffic that fell outside of specified symmetrical boundaries.

Bound on IPv6, pp. 12 - 16

In an interview with Jim Bound, who is the IPv6 technical leader for the Compaq (formerly Digital) Unix group, we have created a very broad and detailed overview of the status and functionality of Iv6. As Bound explains it most of the protocol work is now done, a world wide tunneling application known as the 6Bone is in operation and Unix applications will appear next year along with initial moves to allocate IPv6 addresses. He points out that NAT address translation is a band aide approach that makes the use of IPsec impossible as well as the kind of QoS applications needed for the success of the internet telephony gateway market. Mobile roaming from laptops back into corporate nets also won't work with NAT.

He believes that the urgency for corporations to use these applications will compel them to start to implement IPv6 in their networks next year. While they will have to renumber in transitioning to IPv6, it has been designed to make the process easy by comparison to IPv4. IP numbering allocation has been established in such a way that, by means of top level, middle and local aggregators, routing will be simplified and flaps unable to spill over their aggregated boundaries. Unfortunately the vast number of IPv6 address numbers available will not impact the hierarchical delegation of the aggregators and we may see disputes over number assignments continue.

Blumenthal on Building BBN Network, pp. 17 - 21

We interview Steve Blumenthal, Senior Vice President of Network Engineering at GTE Internetworking. He explains the steps that BBN undertook to integrate BARRnet and SURANet with its NEARnet core between 1995 and 1996. Critical to this process was a need to standardize a myriad of systems from customer care and billing, to network operations, maintenance and provisioning across multiple organizations while maintaining the rapid growth of all three. Lessons learned here are clearly relevant to MindSpring, Verio, WorldCom and others acquiring existing ISPs. Blumenthal also talks about GTEI's provisioning of its 24 strands of Qwest fiber. GTE is using Nortel OC 192 SONET optoelectronics to light up the fiber. The Nortel equipment supports 16 DWDM lambdas on each strand. GTEI is beginning to provision OC3 and OC12 circuits from the fiber and are offering ATM and frame relay services as well.

In order to provision high bandwidth circuits for sale to other carriers GTEI has installed SONET throughout. Blumenthal covers other aspects of GTE's planned expenditures of five to six hundred million dollars a year for network upgrades for the next several years. He concludes with and explanation of GETI's provisioning of America Online and its global provisioning capability as the result of its alliance with Equant, the commercial arm of SITA.

Does IANA Do an Open Process?, pp. 22, 24

Jon Postel and almost half the country code TLD administrators got together in early August and tried to create among themselves a group that would become the Names Council and speak for the commercial registries as well. We show that primarily behind this plan are ISOC and the country code administrators who, while charging lucrative fees, pledge to keep their organization beholden to Jon Postel and to support his IANA draft by laws. We list the fees charged by those countries requiring no operational presence for the administration of their domains.

We also note the conclusion of Roeland Meyer on August 22 was even more striking: "Actually, you make a real strong point. If the IANA folk have a discussion list, it is certainly closed to me, as well as, most that I know. It could be argued, by Postel and Co, that they don't have time for such silliness. However, he's not available to anyone else I know either. If he's doing something, and I'm sure he is, why do we all find out after the fact? That kind of blows his credibility vis-a-vis open processes. On reviewing the latest Draft-Postel, this is kind of underlined (during my feedback to the NSI review, which Dan and I both did). Jon Postel doesn't "do" open processes.

cont'd from p. 22
World Wide Alliance of TLD Registries] would like to make the following announcement: The signatories are pleased to note the intention to establish a new body to administer Internet functions which will continue to uphold the neutrality and high quality of services now given by IANA.

COOK: Neutrality? How many of these administrators who have been enfranchised by Jon Postel with their exclusive rights to administer and charge for domain names can claim not to be "friends" of Jon? I doubt that they will match strict neutrality in any sense.

The web site says: With the following provisions, signatories to this announcement are prepared to work within the framework proposed in the Bylaws drafted by Jon Postel and released on 17 July 1998.

Signatories would like to point out that TLDs currently comprise 7 generic TLDs or gTLDs and over 200 country code TLDs or ccTLDs. In this context, and in keeping with the spirit of the proposals for the address and protocol numbering support organizations (and related councils) the signatories would propose that the World Wide Alliance of TLDs, which is open to all TLDs, be considered the Names Support Organization and be responsible for nominating the Names Council.

COOK: So the 2.5 million holders of .com get the same representation as the Pitcarn islands? Tommy Ho who is registrar for America Samoa and Bhutan gets two votes and .com gets one? These folk are prepared to work within *Postel's* by laws. Of course they owe their fiefs to Jon. It would go without saying that they are not prepared to give him any grief.

ISOC has been flaming about lack of competition for .com in a situation where there are plenty of resellers for .com and where as we pointed out earlier Tabnet built a business that just sold for 45 million dollars by reselling .com. Why doesn't ISOC demand country code competition. Perhaps if there were competition for France .fr, it would lower the $400 a year cost of a .fr domain? And charging $400 or more a year for Congo and Rwanda domain names grossly discriminates against citizens of those countries given their low per capita income. Jon Postel and ISOC should think about what kind of history they are writing for themselves.

Events after August 10th

According to an August 15 posting by Richard Sexton: They announced in Singapore that this [speaking for gTLDs as well as country code TLDs] was not their intention and was just bad wording.

Then on August 18 the group made a fresh announcement: The World Wide Alliance of TLD Registries (WWTLD) announced at the opening plenary of the Asia-Pacific International Forum on the White Paper (AP-IFWP) that over 100 TLD registries had become signatories to its initial position paper <www.canarie.ca/tld. WWTLD is to be established in co-operation with the regional ccTLD groups: CENTR <www.ripe.net/centr and APTLD <http://www.apng.org/apcctld/

Editor's Note: Due to space limitations we must truncate this article at this point in our hardcopy edition. The electronic edition contains the complete article. Missing from hardcopy is a critique of the web site. A listing of most of the 90 signatories with the prices charged and comments on the commercial nature of the signatories.

The COOK Report on Internet
COOK Network Consultants
431 Greenway Ave.
Ewing, NJ 08618