



# Packet Design Has Unique Research Role

## Seeks Improved Routing to Offset Problems Inherent in MPLS

### Judy Estrin, Kathie Nichols, Van Jacobson Want Cost Benefits of Convergence By Leveraging Strengths of TCP/IP Architecture

**Editor's Note:** Before coming to Packet Design and Cisco, Kathie Nichols was at Bay Networks for about 14 months in 1997 and 1998. Before that she held various industrial research positions and was at a start-up company called Com 21, doing IP over cable modems. With Brian Carpenter she is co chair of the Diffserv working group. At Cisco in 1999/2000 she was Director of the Advanced Internet Architecture Group. In the spring of 2000 she left Cisco to co-found Packet Design with Judy Estrin, Bill Carrico and Van Jacobson.

Since we had interviewed Kathie about diffserv almost a year ago, we asked about the progress of diffserv during the last year. She said "diffserv has gotten the mind share and I believe also the Silicon share in the forwarding path. I think that in that sense we have been successful. The worst development is that while we got started with a pragmatic approach, I've seen now a trend of people who argued about ATM quality of service and Integrated Services for years coming over and trying to muck up diffserv by mixing up control plane stuff into the forwarding path. Hopefully we will survive that. Diffserv now is trying to move into implementation. We have been suggesting that quality of service first be looked at on a per-domain basis and we had been telling people this is where we need to stop for now. Go out and implement and return with the experience you gain."

Judy Estrin has been called a serial entrepreneur. She has been involved with the Internet since she did her master's at Stanford University with Vint Cerf during the time he was developing a TCP. Professor Deborah Estrin is Judy Estrin's younger sister. Judy was involved in the beginnings of the local area network industry with Zilog. She and her husband Bill Carrico went on to found Bridge Communications which was one of the first LAN internetwork players and which shipped the first commer-

cial router in 1983. In 1988 they founded a company called NCD which became the leading supplier of X terminals. In 1995 they founded a company called Precept which was one of the first players in the video-streaming arena. Precept was acquired by Cisco and 1988 and, at that point, Judy became Cisco's CTO for a period of two years.

**COOK Report:** It's been reported that you felt a bit constrained in the CTO role.

**Estrin:** Frankly, it's hard to go from being a CEO to a CTO. The CTO is generally more of a matrix role and as the time went by Bill and I found that we wanted to return to building something. There were a lot of positive things that I learned at Cisco as it went through an incredible spurt of growth. One of the more positive things that did happen was I had the occasion to work with Kathie and Van Jacobson. Both Van and Kathie Nichols came to work for me at Cisco about six months after I joined.

When Bill and I were deciding what we wanted to do after Cisco, we knew that we did not want to start up yet another product focused company. This was partly because we did not want to be in the position of starting something new and then after a year, wind up with someone making us an offer that we couldn't refuse and then we'd get acquired yet again. So we were looking for something a little different to do. The classic thing that a lot of entrepreneurs do at this stage is become Angel investors. But we really enjoy building new technology teams. We were looking to do something different. In the past year-and-a-half I had spent a lot of time talking to Van and Kathie and others in the Internet community and learning about a set of problems and opportunities that existed. In doing so I became more and more aware of a dilemma that exists in the present day Internet.

Volume IX, No. 10, January 2001  
ISSN 1071 - 6327

### Research As a Way of Approaching Problem Solving

In the early days of the Internet a lot of the hard thinking went on in the research labs and universities. This work is not getting funded any more because the research world believes that the Internet has already "happened" and now is in the hands of Industry. Industry tends to look at things in different ways. The problem here, especially since industry is working on Internet time and today's stock market drives people to decisions that optimize short-term investment, is we are at a state in the Internet industry where people tend to focus on products that fill short-term gaps. Now it is OK to fill short-term gaps as long as you don't ignore work on long-term scalable solutions. Coupled with these interests is a concern that we are risking losing the soul of the Internet — namely the basic IP architecture that got us where we are today. This is happening because there are too many forces going on that are pulling us in other directions.

We are merging the worlds of telephony and

On the Inside:	
Packet Design	pp 1 - 6
Exchange Points	pp. 7 - 15
ENUM Agreement	pp. 16 - 17
IS-IS Bug	pp. 18 - 19
ICANN Authority	pp. 19 - 22
Executive Summary	pp. 23-24

data — namely the Internet. As we merge telecommunications and data communications, if we're not careful we will wind up with the worst of both worlds rather than the best of both. If you think about the guiding philosophy of what drove Bill and I and Van and Kathie to start this company, it was a shared interest in building something that would help move forward the best of both worlds.

**COOK Report:** When you look at the definition of the Internet research encompassed by things like the vBNS, Next-Generation Internet, and Internet 2, how would you compare what they are trying to do with what you feel needs to be done?

**Estrin:** Research means different things to different people. Sometimes research means looking ten years out. On the other hand, to us at Packet Design, research is a way of approaching problem-solving. Typical product development means that you have an idea for a product, and then go out and design and implement the product. There is the strategic nature of what product to develop, but when you choose a product to bring to market, what you do at that point becomes very tactical. Everything you do is typically iterative and very focused on the goal of implementation.

The fact of the matter is that there are certain problems that are not solved by introducing yet another new box and yet another iteration of a new and faster processor. Research to me is a way of applying measurement, analysis, and algorithmic thinking to solving problems. In other words we need to approach problems in a slightly different way that leads one sometimes into solving things as opposed to generating a completely new idea. I would use the following analogy if you have a crack in a wall, you can paint it over but, if the crack is structural, in three months it will come back. If you really want to fix it, you must bring in a team that can do structural analysis and understand how the wall was put together. If they are really smart they will team up with implementation folks and solve the problem not by knocking down the wall and rebuilding it, but rather by putting in a shim somewhere in order to solve the structural problem and then repainting over it.

At Packet Design, when we're talking about research thinking, is not just about what the world will look like in ten years. Instead it is approaching problems with the type of problem solving that involves analyzing what is really going on, how it works, and then what you can do to make it work even better.

**Nichols:** We want to think like researchers and to have a vision of where we would like to be in five or ten years but we also want to

do practical things along the way. Therefore if you go back to Judy's house model while we have a general architectural vision of how we want to remodel the house, we also know that we're going to get there in bits and pieces — one step at a time. We want to be able to live in it a while we remodel it.

**COOK Report:** If you are talking about the big picture of convergence, then, from the context of what you said, you must have a concrete idea of how you want to get it properly done.

**Nichols:** Yes and even though we're not waiting five or ten years to turn something out we want to keep an understanding of our end goals in mind.

**Estrin:** Kathie is building a team of people that come from that same research oriented world and we are complimenting Kathie's team with development teams. We couple both groups with customers so that on each one of our projects we engage with sets of customers to make sure that what we are doing is grounded.

We are not trying to create a think tank. We really want to keep the company relevant to what is happening in the Internet. We believe that where a lot of laboratories go wrong is they get too much into pure research and not enough into applied problem-solving. Or what is especially true in big companies, they become disconnected from the customer. The reason for this is that corporate laboratories often do not talk to the customer but rather through too many layers. When we invite customers in here to talk, we often have the entire company sitting there listening in order that people are connecting directly and with out a lot of filters between the problems that people are seeing and the folks here who would like to solve those problems.

**COOK Report:** There was a recent and very fascinating report on Tony Li and his new company called Procket. Are you here in part to be a resource for people like Tony?

**Estrin:** I think Tony and Procket are building a router. Tony would like to build the next Juniper. Tony is very smart. But he has taken the path of developing a new product.. On the other hand Packet Design is a technology company. We now have six ongoing projects. The projects have a measurement and analysis part which overlaps all of them. We have dedicated prototype teams to each. We will rapid prototype all of them within Packet Design, and then we will either license the results, or we will spin out companies to productize our solutions.

The reason that we chose to do this this way is that I have found that when you build a

product company, you are able to innovate, in the first generation but that it gets harder and harder to continue to innovate over time. The reason for this difficulty is that people get tied up with previous iterations of the product that must be supported. What we wanted to do is build a kind of perpetual startup. We have a set of technology problems. Hopefully with our influence on the things we do and after working with service providers and vendors, three years from now we can look back and say we made a difference in the further evolution of the Internet. And by then we will be on another set of projects that will relate to new and different problems. We as a team want to be sure that we are building something for the long term and that what we are building is a sustainable model. This is why we chose to do Packet Design as we did rather than saying that we are going to go off and build a company that would just do routing software.

**COOK Report:** Say that I am in the research and development arm of a company like SBC or Cisco. Are you there so that I can come to you on a possible research task that interests me and find out if it also interests you and whether some joint work between us might, as a result, be possible?

## On the Importance of a Technology Passion

**Estrin:** The answer to that question is no and yes. We're really not a consulting house or an incubation facility. Our primary goal relates to our vision of the Internet and the set of problems that impact the possibility of achieving that vision. We intend to prioritize these problems and to select from those that fall into high priority categories ones on which we think our work will have a significant impact. And as I have said, we will spin out or license the results. We will partner with large companies like Cisco, with service providers, and others for a variety of reasons. One might be the exchange of ideas about what problems there are to be solved. Another is when we might license or sell a spin out company to a larger company. There could be other areas in which we would collaborate with such companies. After all since we're not building a router, if one of the problems we saw was that a change was needed in the forwarding path, we might collaborate with Cisco or Juniper to get the feature implemented.

Our goal would be to make sure that router forwarding paths had the right capabilities within them — and, in doing so, to make sure that the Internet can grow in the way in which it needs to. So we will collaborate and we will partner because we know we're not the only people with good ideas. But I also don't want to give the impression that we are here to incubate other people's ideas be-

cause we have some very specific opinions of our own on the things that we want to do.

The point that I wanted to make about communicating and opening up to the outside world is that we must take care not to become insulated from that world as we pursue our objectives. Therefore we're being very pro-active in making sure that we get lots of information about people's problems. We are paying close attention to the ways in which other people are proposing solutions. We do this in order to be sure that our products will mesh with the current needs of the Internet. The projects that we are working on are all internally generated. The people working directly on solutions are all Packet Design employees.

When we prototype something we will either have a wholly owned subsidiary for which we recruit an operational team. Or we will license it. It will all depend on the size of the idea and whether we think this is a single product idea that someone will acquire quickly. Or whether we think it is the basis of a company that might want to stand alone and eventually go public. We will decide all this on a case-by-case basis.

**Nichols:** As you said we're very opinionated about the shape of the Internet and the shape of solutions to those problems. While we want everyone to know that we are open to hearing about problems and their ideas for solutions, we also want them to know that we intend to incorporate their input and formulate solutions according to our way of thinking. We don't want to mislead anyone who may tell us: oh, you ought to solve it this way into thinking that's exactly what we will do.

**Estrin:** One of the things that I think has contributed to Cisco's success, is that Cisco will very publicly say that a key part of their culture is that they have no technology religion. What they mean by that is they are not going to be so religious about IP that there won't ship ATM if the customer would really rather have it. Or vice-versa. And when you are company that is Cisco's size or in even Juniper's size, you can understand saying that I need to make sure I'm providing my customers with what they need and that I am not going to get fanatical or religious about a certain approach.

What often happens though is that "no technology religion" can get mapped into "no passion for a specific technology solution". There are certain hard problems that, unless you have the passion to keep pursuing a solution, you will not overcome the obstacles that need to be overcome. If you don't have a passion for something, when you hit a wall, you will simply go off in another direction.

**COOK Report:** So what are your particular

passions?

## Why the Internet Has Succeeded

**Estrin:** I think that the fundamental issue at a very high level is that if you look at the design of IP there were a number of things about it that brought us to the point where we are today. It is open and available to everyone, encouraging rapid, competitive innovation. It is application-independent, requiring no proprietary application-layer gateways.

And probably most critical, its designers made three key decisions about its distributed architecture. First, because IP provides for separation of the control plane and the forwarding path, major advances in these two areas could be made independently: forwarding speeds could be pushed to the limit in silicon, while routing control functionality could be improved concurrently in software. Second, services are placed at the edges of the network rather than integrated into the network itself; this allows services to evolve without impacting the network and keeps complexity out of the network core. Third, and perhaps most important, is the fact that IP was designed to have globally known addresses. What this allows you to do is to build a distributed architecture in which each device in the network has all the information that it needs to make decisions in processing each packet. This allows for the distribution of work throughout the nodes, both providing redundancy and improving scalability.

The analogy that I like to draw here is that if you think about organizational structures today, it is commonly known that the way organizations of people scale is that you empower people with information so that they, acting independently, can make correct decisions and can do so at the correct level. We have gone beyond the very hierarchical organization where only the people at the top have the necessary information for telling the drones below them what to do. Everyone knows that this doesn't scale. So if you think of the distributed architecture of the Internet, you will realize that it does exactly that. Namely it makes sure that each router and each node in the network has the information necessary to make a decision.

The distributed nature of IP's architecture makes it very different from the public telephone network, or PSTN, whose design is centralized and more complex. Telephony networks, which are based on connections, or circuits, work more like organizations with rigid hierarchies, where individuals (in this case network switches) are told what to do by a central controller. Circuit-based network architectures are inherently limited in

scalability because new connections ("strings") must be set up, and consequently managed and accounted for, every time a new element is added.

## The Problem of MPLS Arrose Out of a Vacuum of Routing Expertise

Now in this context there are some trends that concern us greatly for example what is going on with MPLS. With MPLS the problem is that what started out as a very specific solution to mapping IP and ATM has now gotten out of control and is being presented as the solution to all sorts of problems. In this case MPLS does violate at least one, if not more, of the critical elements that I talked about: IP's globally known addresses. Several years ago Van gave a talk that he called "Clouds verses Strings". The idea was that while the Internet World looks at clouds the circuit-switched world looks at strings or connections. MPLS is putting strings in those clouds. While it is OK to put clouds over pieces of strings, your end-to-end architecture needs to be clouds. If you start imposing too many strings, you begin to break the basic scalability of that architecture.

The convergence of data, voice and video we are seeing today is driven by the dramatic increase in data traffic now being pushed across the PSTN infrastructure. The traffic patterns associated with new data applications are very different from those of phone conversations. Yes, the Internet needs to maintain the manageability of the telephony world -but not at the expense of scalability. MPLS, a string-oriented technology, was developed to solve a point problem: integrating local IP and ATM environments. It works well for that use, but its proponents have positioned it as a panacea for all sorts of other problems. At first glance, MPLS seems like the perfect answer to a converged Internet. But it's really just a quick fix. Because its architecture is based on strings rather than clouds, it has all the disadvantages of strings and, in the long run, it creates more problems than it solves.

**COOK Report:** In other words like Mike O'Dell are you building a network core of fully meshed permanent virtual circuits?

**Estrin:** That is one way it is done. I am not saying that all service providers should just do away with MPLS immediately. What I am saying is that MPLS should be regarded as a short-term solution to the problems of traffic engineering. If you ask people why they run MPLS, you will find that they do so for the same reasons that they adopted ATM.

MPLS satisfied the need for a level of traffic engineering and traffic management and determinism in networks. Today people don't feel comfortable about having that with only pure IP and with clouds. But we believe that anything that can be done with MPLS can be done with a cloud and with an IP-based architecture without breaking a fundamental design criterion of IP. Some of this takes some work but it can be done.

Unfortunately in the early nineties, the use of routing as the primary basis for data networks gave way to Ethernet switching — a technology that was faster, cheaper and easier to understand. Most companies and universities turned their research and development efforts away from routing during this period. MPLS arose out of a research vacuum, a void of routing expertise. But while circuit-switching may work fine in a relatively static local network, a connection-oriented approach that can't scale to hundreds of nodes, much less millions, cannot be applied to create an effective long-term end-to-end solution. And the Internet is the least static and least localized network of all time.

**COOK Report:** But aren't people saying that if you don't do MPLS like kinds of things, you will end up with a backbone having so many flows as to make your backbone become so large and unwieldy as to be unmanageable? So is solving the issue of backbones size and manageability one of the tasks that you believe you can take on? And what would be the downside of not doing this?

## Co-mingling the Control and Forwarding Paths

**Nichols:** We absolutely do think that we can look at traffic management and traffic engineering from an IP perspective. To get back to what Judy said early on, we want to keep that forwarding path simple. One of the problems with MPLS is that it's leaking some of the control plane into the forwarding path by putting its label into the packet (or forwarding path). Consequently when you're trying to move the packet through the network, you have a control label out there someplace in the middle of the network. If a node goes down, you have to go all the way back to the edge in order to pick up the circuit that it belonged to. Whereas with IP, if one route doesn't exist, you go to another one.

**COOK Report:** In other words because the envelope containing the address for the packets is buried in the middle and you cannot recapture it or recreate without starting over.

**Nichols:** If you really have an MPLS type of structure, this is correct. Now there are a couple of things we can do. With regard to

traffic engineering and traffic management, we think that we can do this in an IP kind of way. Admittedly this will take some work. We have also looked at something that has been somewhat confusing to us all along. Namely, even if we were to accept that doing circuits is the right thing, why you need that tag in the packet since the tag is created from information that is already in the packet?

The forwarding path of IP has been evolving very nicely. We have fast classifiers. We have fast packet handling. Since we have all this great stuff, we should just use it. You can classify those packets and can get that same information anywhere in the network.

**Estrin:** Let me say again that we understand why MPLS happened. The reason that it happened was that work on routing stopped several years ago and that except for a very small group, there has not been much design and architecture work done in the routing area. Most of the work in routing in the last couple of years is been along the lines of "oh my God how do I keep up with all this growth?" It has been the putting of small configuration options into IS-IS as a work around to a certain crash that happened in the Internet. Instead of just continuing the iterative thinking, it is time to take a step back. We must acknowledge a widespread need for taking routing to its next step. Now while all of this has been happening, there has been an enormous amount of energy going into switching.

**COOK Report:** Noel Ciappa, as one of the first generation routing architects has been extremely critical of BGP in recent years. Do you share some of his criticisms?

**Estrin:** We may see some of the same problems. When we look at the protocols today we find that some of them have lots of good things and a few bad things and that some of them have few good things and lots of bad things. There is one important difference regarding our approach. As we pointed out, we are trying to take into account the need to live in the house as we remodel it. Therefore we do not propose throwing everything out and starting over from scratch unless we get to some point where we just throw up our hands and say: "impossible." We believe that we have the beginnings of the answers. And yes while we do need to prove them to ourselves, we do think they will work. However we are not yet ready to disclose them in this conversation.

**COOK Report:** What you are doing then presumably is communicating enough with outsiders so that you don't develop your ideas in an ideological isolation?

**Estrin:** We have spoken to several service providers and continue to do so. We gener-

ally start by asking them to tell us what their problems are. We try to get them to start off the discussion. While we do admit to having a passion we also engage in reality checks.

**Nichols:** We have some strong beliefs about how you go about solving problems. And I think you can also say that we have an architectural bias. We have our bias in building the Internet. However, we do not have a bias about precisely what the most important thing to do right now is. In the small amount of time that we have been in existence it has been a great experience for us when we do talk to people because exchanging ideas fires us up. It allows us to continually rethink what we want to solve first and how we will deal with it.

The point that I wanted to make earlier is that while we think MPLS in the control plane is not the right answer, if we were to talk to enough people and discover after all that it is the right answer, we can cope with that, because we are pragmatic, but we still don't think it should be done in a way to complicate the forwarding path.

## Doing Something About Routing

As Judy said, routing has not been looked at in quite some time. We therefore said that the first thing that we have to do is see what kind of shape it's in. What we have indeed heard from many the people we've talked to in the industry is: "do something about routing."

**COOK Report:** What are the kinds of things that can be done about routing?

**Nichols:** that's a good question to ask because we gave a presentation at NANOG 20 in Washington about one area that we have been investigating. We set out to find why routing convergence times don't work any better than they do.

**Estrin:** Before putting together the NANOG talk, we went to the vendors whose equipment we tested and showed them our results to make sure that they understood them. I think that our work will help them to focus in on some areas a bit better. And the reason for the NANOG talk is that we want service providers to understand what can be done if you take a step back and look at some of these things. We also want to encourage them to work with us so that we can gather more data and do more analysis to understand even better what is going on. A lot of what we've done so far is working with tools that allow us to build virtual topologies instead of physical ones.

The NANOG paper is titled 'Toward

Milisecond IGP Convergence.” See <http://packetdesign.com/Docs/isis.pdf> > It is by Cengiz Alaettinoglu, Van Jacobson, and Haobo Yu. It suggests that sub-second re-route times would give increased network reliability; support for multi-service traffic (e.g., VoIP), and lower cost/complexity when compared to layer two protection schemes like SONET. Since current IP re-route times are typically in the tens of seconds, we need to do better. There are two choices: replace IP routing with something else like MPLS fast failure recovery or figure out what’s wrong with IP routing and fix it – by now you probably have figured out that we believe it should be the later.

**Editor’s Note:** They have reached some interesting conclusions. The following seven paragraphs are direct quotes from the slides of the presentation that identifies and recommends significant changes to the is-is spec. They show in a most convincing manner how these changes should drastically reduce routing convergence time. The 21 slide presentation includes nine graphics representing results of detection experimentation, LSP propagation, and SPF measurement and calculation. We encourage readers to grab the 250 kilobyte PDF file from the Packet Design website. We include the quotations to whet their appetites

“The IS-IS spec allows for two link state change detection mechanisms: link-level notification and peer-peer Hello packets. Link level detection should be fastest but it’s not always possible (e.g., switched Ethernet) and seems to be inconsistently implemented by vendors. Spec says that adjacent routers should send Hello packets to each other at a fixed interval (default 10 sec., minimum 1 sec.) and declare the adjacency lost if no Hellos received for three intervals. This works for any interconnect but constrains the repair time to be at least 3 seconds (3 times the Hello interval).”

“Since the protocol’s ultimate limit on the Hello interval is set the bandwidth used by Hello packets, extending the spec to allow sub-second intervals would allow sub-second detection on almost all links. With this change, detection time is limited by the physical constraints of the transient error spectrum on a particular link. For example, a link that takes 30ms noise hits should have at least a 30ms Hello interval.”

“For either detection method there are network-wide stability issues if routing tries to follow rapid link transients (i.e., a link that goes down and up several times a second). The usual way of dealing with this is to treat “bad news” differently from “good news” so routing is quick to find an alternate path on any failure but slow to switch back when the link comes up. The current IS-IS spec treats bad news and good news the same but

it should be trivial to change the detection spec to allow different filtering constants for “down” and “up” state changes.”

“A link state packet is generated at the point of detection then flooded, unmodified, through the network. It should propagate at near the speed of light plus one store-and-forward delay per hop. So in theory LSP propagation should make a negligible contribution to the re-route time. Theory doesn’t often resemble reality... [Editor: graphics on LSP propagation follow:]”

“LSP propagation explanation: Since the SPF calculation can take a significant amount of time (see next section), commercial router implementations impose a limit on how frequently the calculation can be done. In some implementations this limit is fixed (5 seconds) and in some it is changeable but with a granularity of 1 second. This limit essentially adds to the propagation time. On at least one implementation, if the SPF limit is set to zero, the SPF calculation is done before the changed LSPs are flooded which degrades the propagation time from  $O(\text{speed of light})$  to  $O(\text{diameter} \times \text{SPFtime})$ . To prevent this, the spec might be amended to explicitly state that LSP flooding is “higher priority” than SPF calculation or the point might become moot if the SPF calculation time is improved (next section).”

“After any link state change, each node has to do an SPF calculation to compute the new topology. Even with high-end platforms (Cisco 7200, Juniper M40) this calculation can take a lot of time (seconds) and has poor scaling properties ( $n \log n$  to  $n^2$ ). This has a serious impact: on convergence (because the SPF calculation is in series with the LSP propagation).[and] on overall network stability (because of router CPU saturation).”

“The Dijkstra SPF algorithm” used to compute changes to SPF trees (route forwarding tables) “is almost 40 years old. More recent algorithms can compute changes to SPF trees in time proportional to  $\log n$  rather than  $n \log n$ . This allows a net to scale up to virtually any size while bringing the calculation time down from seconds to microseconds. Consequently stable, robust IP re-routing that works at the network’s propagation rate (the theoretical maximum for any re-routing scheme) is both possible and achievable. To get there we have to (in rough priority order): (1.) switch to a modern algorithm for SPF calculation, (2) make the granularity of the hello timer milliseconds rather than seconds, and (3) allow different detection filter constants for link up and down events.”

**Estrin:** The problems that we have with IGP’s now are really due to the fact that the vendors designed and implemented them a

good number of years ago. They have iterated upon them to fix problems that turned up in the field. Perhaps it’s time to step back and say is there a way to leapfrog this and do something better because we’re no longer dealing with just file transfer. We will be dealing with increasingly critical applications. We may speculate all we want about why the vendors haven’t looked this. The fact of the matter is that as we look into the future a year or two, we believe that this is going to be an increasingly important problem. We are in a position where we can try develop some new approaches.

**Nichols:** I think that this gets back to what we’ve said all along. Many people have become convinced that routing is a very mysterious thing. Because you have just a few people who know what to do about it, it becomes black magic instead of something that you can look at and improve. This is kind of interesting because people have looked at the forwarding path and done I think really great things. For while this area had been considered unimprovable but then suddenly a few years ago there was a vast improvement in the ability to do IP look ups. We are trying to do the same thing for routing. We’re trying to say look you can apply some science to this.

**COOK Report:** If you can get in changes like this accomplished, doing so should have a very powerful impact. But let me also ask whether there is any reason to assume that you might also be able to work with CAIDA on doing this?

**Nichols:** As far as I know they are not doing anything that is exactly like what we’re doing. We are pursuing our own efforts to get measurements. We think this is somewhat better for us because we then have a direct relationship with whom we are getting the data from and with whom we would sign a nondisclosure.

**Estrin:** When you ask what things need to be improved, you’ll find out that there are lots of things. When you go to different service providers with different configurations, they will point to lots of things that are difficult about IP. Some of them have to do with provisioning and management. Some with traffic engineering. Certainly the network convergence time issue that was just mentioned is very significant.

## Preserving the Advantages of the Internet’s Distributed Architecture

One of the things that I wanted to say earlier was that while routing is a big area that we’re working on, it is just one of the things that

we're working on. These include some other things in the infrastructure area, some things in security, and some things in web acceleration.

Something that is very very important to us is the ability to get raw data and to do so in ways that our solutions are not to jaded by any level of indirection on the part of the people who gathered the data. We reached out to people for the collections. We have looked at reports. And all of this helps us form hypotheses. But because so much of what we're doing is really based on looking at what is actually going on, what is really critical to us is to make sure that when we measure and we capture we also own the data so that we can go back again and look at in a different way if our first pass turns out to contain something of which we are suspicious. We consider data collection a critical core competency for what we're doing and therefore we will likely want to be responsible for doing all of it ourselves.

**COOK Report:** Is part of the problem that you're getting at in being critical of the wide extent of the application of MPLS the fact that you believe that, as the use of MPLS expands, the cost of running networks using it increases in comparison to the cost of being able to run an Internet without it?

**Estrin:** Absolutely. The reason that we feel so strongly about this is that we believe that it is the properties of IP which can take us directly to the type of Internet that yields the best scaling characteristics at the most favorable cost from a manageability perspective. Now what the telephony world does very well is to offer us some best practices in the areas of manageability and billing and accountability. We need to map these onto the Internet world. But you don't achieve this by making the Internet infrastructure look like the telephony infrastructure. You need to figure out how to do all those things within the confines of clouds.

**COOK Report:** How about the attractiveness for you of gigabit and ten gigabit Ethernet over fiber? If you succeed in doing what you want to, you will leverage tremendously the already preeminent cost effectiveness of doing Ethernet won't you?

**Estrin:** Absolutely. All of this work is so especially important right now because of the convergence of telephony and data. It is also important because of the introduction of optics into the equation because we are building more and more bandwidth and pushing into faster and faster systems. There are two architectural approaches when you are thinking of optics. There is the traditional world which says that you build faster and faster centralized switches which is what telephony switches look like. Or there is the

notion of you taking advantage of what got us to here which is an inherently distributed architecture. But there are hard parts about being distributed including problems that must be solved. I think that there is no question that if we do what we set out to do and do it correctly, it will help enable and be applicable in a wide range of areas that the Internet needs to go into.

This is really about the level of detail that we're comfortable in going into with you today. Because after certain point in talking about problem it is hard to continue further without describing the solution and it will be a few more months yet before we are comfortable in doing that. In addition to my mention of web acceleration and the security area, let's just say for now that we are working on some ideas in the area of mobility in other words wireless.

**COOK Report:** OK. Having said all that, would you be willing to conclude our discussion today with some comments on Bill St. Arnaud's ideas about an optical border gate way protocol?

## Assessing OBG

**Nichols:** I've read your interview on it and discussed it a bit with Van Jacobson. We do think it's great that someone in a research network is looking at this problem and trying things out. (We're glad that someone doesn't say the answer is MPLS J ) One of the things that Van pointed out, however, is that he is not sure that the numbers work out. He says that there are about ten gigabytes in a single color or lambda and that at last count there are about 500 million systems connected to the Internet. You have about ten very large providers doing global backbone service where each backbone will have about a thousand colors or wavelengths. The notion was how will you make this per color route go through the skinny backbones without running into problems.

**COOK Report:** But I thought St Arnaud's idea was to be able to buy a single wavelength and to send it to another point with which you wish to communicate. The whole purpose in doing so was to avoid having to transit the large overcrowded backbones.

**Nichols:** True, but how does that work in terms of building an infrastructure? We're not opposed to his ideas, but as we said before, our plans are more to work on fixing the house as we live in it. I think what St Arnaud proposes requires an assumption that the fundamental structure of the infrastructure will change. Maybe this is also a good illustration of where real research differs from what we are doing. In true research you may do something risky that is based on a belief that the infrastructure will change in

ways that don't seem likely at the present time. It looks like he is working on that kind of a basis which would be too risky for us.

**Estrin:** Just to summarize, and as I do so please understand that no one here has spent a lot of time studying OBG, therefore all we can do is give you an initial reaction. We are glad to see people looking at this area and figuring out how it needs to evolve. Our initial thought is that this may not be the answer. Over time we will probably look at it more and we are certainly not suggesting that we know enough yet to say that it will never work.

I think that over time when we look at the Internet and optics we may see a number of approaches being floated. Perhaps there will be an answer or of some combination of them. I think what you're hearing is that when we saw this work, our reaction was not: "oh! The problem is solved."

**COOK Report:** To what extent do you think that there may be political issues that may cause it difficulty? For example things like router companies being slowed to implement the OBG extensions in their code or backbone operators not cooperating?

**Estrin:** This is precisely what Kathie meant when she remarked about the risk involved in any technology that requires a major change in infrastructure. This is something I learned in spades at Precept because we built a company, on the basis of an assumption that multicast would be widely deployed. And guess what? It wasn't. So the one thing that we've learned from experience is that if you're designing a solution to a problem where you assume that all infrastructure involved needs to change and do so at the same time, you have a very big problem on your hands.

Look for example at IPv6. One of things that has saved IPv6 from being a complete no starter was that there was very considerable thought put into coexistence and compatibility issues with IPv4. You can never say to the Internet: "OK. Switch." Therefore I think that the author of any solution to a problem today has think through the issue of backward compatibility very carefully. That doesn't mean you have to limit the span of what you're doing, but if it is something unusual, you have to put other things in place to be able to mitigate any compatibility issues. Consequently, I think any approach that expects everybody to adopt something and change their basic way of doing things has the odds against it. One of the problems with MPLS is that while it does not require that everything change, those boxes that it touches will incur very significant and difficult to achieve changes to their code.

# Scaling the Internet via Exchange Points

## Many Players Jumping into Rapidly Expanding Global Market Technology and Business Model Issues as Seen By Equinix

### Editor's Introduction

The Exchange Point portion of the Internet Industry is set to grow dramatically. It is motivated by the need for ISPs to peer with as many other ISPs as possible and at as many locations as possible to cut down on the money they must spend on buying transit from Tier One ISPs. Also exchange points help keep local traffic local and in doing so cut down on the amount of traffic that would flow to otherwise overcrowded long distance Internet backbones.

Exchange Points started out as vertically integrated carrier operations where ISPs would interconnect to a central switching fabric. The NAPs (Network Access Points of Sprint and Ameritech) were early examples as were the MAEs (Metropolitan Area Exchanges) of Metropolitan Fiber Systems (MFS). By 1997 a problem with this carrier owned exchange business model became evident. Traffic grew so fast, that Gigaswitches could not keep up. There weren't any good vendor alternatives to the switches for a long time, with the consequence that early on the MAEs were victims of their own success. Also there was a two-fold organizational conflict of interest. MFS wanted to increase circuit sales, so their focus was not on other operational and facilities aspects of the MAEs, such as making them larger for web hosting. UUNET (like all the major backbones) wanted to drive transit sales. As a consequence they agreed to peer with relatively few other networks.

### New Models Spring Up

In late 1997 the Palo Alto Internet Exchange <<http://www.paix.net/>> pioneered the idea of a neutral exchange point where the business model is finding a secure building with secure power and multiple providers of dark fiber. Tier One ISPs were invited into this neutral model as anchor tenants for a mall. Smaller ISPs then saw it as an opportunity to collocate and peer. In efforts to make the web work better, web hosting was moved in as a major collocation tenant. The neutral model fit well the tendency of ISPs to become collections of horizontally grouped services where an ISP could outsource many of its functions such as email service to a company that specialized in running SMTP servers, and its web services to a hosting specialist.

In 1998 AboveNet opened large exchanges in Virginia, California and New York connecting them with very large bandwidth. At this time AboveNet also bought PAIX. In October 1999 AboveNet opened its European Internet Service Exchange (ISX) facilities in London, Frankfurt and Vienna offering co-location services and common peering connections among ISPs. To support the exchange points, AboveNet added an STM-4 line Trans-Atlantic link to its network. Its aggregate Trans-Atlantic capacity now exceeds 750 Mbps." AboveNet advertises peering with 420 ISPs and seven Trans-Pacific links. /" <http://www.above.net/>

Collections of data about exchange points can be found at Exchange Point Net <http://www.ep.net/> and at Telegeography's Internet Exchange Points directory <http://www.telegeography.com/ix/> which has a somewhat dated listing of "over 200 existing and planned Internet exchange points. City location, URL, and where available, number of co-located ISPs and average traffic load are shown for mid-October 1999

A valuable new overview November 2000 article "Carrier Hotels: Click and Mortar" by Sean Buckley Staff Editor of *Telecommunications* is found at [http://www.telecommagazine.com/issues/200011/tcs/carrier\\_hotels.html](http://www.telecommagazine.com/issues/200011/tcs/carrier_hotels.html). UUNET, ATT and Cable and Wireless run vertically integrated exchange points. On examination we found that the definition of exchange point for these carriers appears to be quite malleable being what many would simply call a network node. For example. C&W's European points are found at [http://www.ecrc.de/netzwerk/netzwerk\\_ncp.html](http://www.ecrc.de/netzwerk/netzwerk_ncp.html) and some idea of its planned 84 network nodes is to be found at <http://www.cablewireless.com/default.asp?PageID=43>.

Certainly the neutral model pioneered in 1997 by PAIX is where the growth is. As Buckley points out "companies such as COLO.COM, CoreLocation, Equinix, Eureka, MFN's PAIX.net, and Switch & Data are expanding into major telecom hubs. Unlike the traditional telecom hotel model where carriers lease physical space from a building owner, neutral COs provide air conditioning, backup DC power, HVAC, dust control and high-level security in addition to real estate."

While an examination of the web sites of these players showed business models in

varying states of elaboration, the predominant emphasis was on the bare bones provision of a facility with fiber, power, security and environmental controls. For example, at [http://www.color.com/english/about\\_us/index.htm](http://www.color.com/english/about_us/index.htm) Colo.com brags: "The Neutral Optical Hub solution for collocation will enable Bandwidth Intensive Businesses, such as ASPs, ISPs and CLECs the choice to access advanced network and facility resources, deploy distributed networks, and deliver content-rich interactive applications and services close to their end-users." It claims 13 open locations in the US. CoreLocation is opening (with its venture partner, the Carlyle Group) exchanges in Chicago, Atlanta, and San Jose. The facilities are very large (multiple hundred thousand square feet) and located in old buildings undergoing renovation. See <http://www.corelocation.com/projects.html>.

Switch and Data Facilities (S&DFC) founded in 1997 advertises at <http://switchanddata.com/> a world wide network of carrier neutral, co location facilities which are mostly very small in size ranging from 10 to 20 thousand square feet. "S&DFC's current customers include ISP's, Competitive Local Exchange Carriers (CLECs), Data CLECs, Digital Subscriber Line Providers, voice processing companies, and corporations needing hot backup sites." Twenty sites are currently opened with another 21 planned for opening in the next 6 months.

According to Buckley: "In any nascent market, each player has its own approach. Major collocation providers differ in size, capital and purpose. At a minimum, companies such as COLO.COM, Layer One, CO Space and Switch & Data provide real estate rack space. Because their facilities cover about 20,000 square feet, they have expanded to several sites in the United States and internationally. COLO.COM's expansion plan calls for at least 40 facilities. Unlike the others, Layer One provides interconnections among backbone providers coming into its facilities and expects to complete 12 neutral COs by the end of the year. PAIX.net, one of the first Internet exchanges, [and now a part of MFN AboveNet] expects to open six new facilities in the United States by the end of the year." [For Layer One see <http://www.layerone.com/home.php3> ].

"While the neutral CO [Co Location] has a bright future—Ovum [the United Kingdom's largest telecommunications consultancy] predicts revenues will be \$55.8

billion by 2005—there are challenges. The demand for neutral CO services is here, but successful companies must execute four elements effectively: capital, connectivity, power and location."

"Because the neutral CO is fundamentally a real-estate issue, entry requires up-front capital. Most neutral and even non-neutral colocation facility owners pre-sell the space, buy the building and build the facility based on presold commitments. . . . Despite a neutral CO's proximity to major fiber routes, there is no guarantee a backbone provider will connect to the facility. While demand will not end soon, there is concern that the colocation industry is heading toward a pseudo space glut with plenty of space but nowhere to connect to the backbones. It is estimated there are 42 national CO providers, with more than 25 million square feet coming on-line next year, a 50 percent increase over current availability."

## Equinix

While the MFN's Paix and AboveNet run multiple neutral exchange points and with roughly three years of experience are certainly well aware of what it takes to keep customers happy, at the close of 2000 Equinix seems to be the industry player with the most detailed knowledge of industry needs and dynamics. Having the original founders of the PAIX, Equinix also has a handful of technical experts with many years experience in the industry. These people are likely to be known by nearly all of Equinix's customers. Consequently they will offer a significant but yet intangible benefit of 'cluefulness' on the calculations of those considering locating in IBX's.

In addition to the gathering of substantial industry expertise Equinix's business model seems to be founded on the conclusion that there will be benefits of scale to the player that can establish a global set of very large exchanges that are populated with a range of players extending well beyond the carriers, ISPs and Web hosting players common to the rest of the industry. With the exception of CoreLocation the size of an Equinix IBX is about ten times that offered by its competitors. Equinix is convinced that, seeded by critical players, their large facilities will each attract a very large number of ISPs, hosting providers, mail providers, storage providers, ASPs and so on. The numbers and diversity of the occupants of the IBX 'mall' the better for the other tenants and the better for Equinix.

The down side of this strategy is that it is very capital intensive with each mega-facility demanding on the order of \$100 million to open. While Equinix has so far raised on the order of \$600 million, its very ambitious global development plans demand brilliant

execution if it is not to stumble. The dangers it faces are laid out on pages 7 and 8 of its June 21, 2000 S1 filing with the SEC. < <http://www.edgar-online.com/bin/edgardoc/DocFrame.pl?doc=A-1101239-00001012870-00-003459&fmt=text&nav=&nav=1&x=81&y=29>>

"In a market that we believe will likely have an increasing number of competitors, we must be able to differentiate ourself from existing providers of space for telecommunications equipment and web hosting companies. We may also face competition from persons seeking to replicate our IBX concept. Our competitors may operate more successfully than we do or form alliances to acquire significant market share. Furthermore, enterprises that have already invested substantial resources in peering arrangements may be reluctant or slow to adopt our approach that may replace, limit or compete with their existing systems. If we are unable to complete our IBX centers in a timely manner, other companies may be able to attract the same customers that we are targeting. Once customers are located in our competitors' facilities, it will be extremely difficult to convince them to relocate to our IBX centers."

"We intend to construct IBX centers outside of the United States and we will commit significant resources to our international sales and marketing activities. Our management has limited experience conducting business outside of the United States and we may not be aware of all the factors that affect our business in foreign jurisdictions. We will be subject to a number of risks associated with international business activities that may increase our costs, lengthen our sales cycles and require significant management attention. These risks include:

- increased costs and expenses related to the leasing of foreign centers;
- difficulty or increased costs of constructing IBX centers in foreign countries;
- difficulty in staffing and managing foreign operations;
- increased expenses associated with marketing services in foreign countries;
- business practices that favor local competition and protectionist laws;
- difficulties associated with enforcing agreements through foreign legal systems;
- general economic and political conditions in international markets;
- potentially adverse tax consequences, including complications and restrictions on the repatriation of earnings;
- currency exchange rate fluctuations;
- unusual or burdensome regulatory requirements or unexpected changes to those requirements;
- tariffs, export controls and other

trade barriers; and

longer accounts receivable payment cycles and difficulties in collecting accounts receivable.

To the extent that our operations are incompatible with, or not economically viable within, any given foreign market, we may not be able to locate an IBX center in that particular foreign jurisdiction."

One thing is certain. In order to continue to scale, those building the Internet must figure better ways to interconnect. Getting an up date on the technical, operational and economic issues involved in interconnection will give a worthwhile picture of who the players are as well as how they are interacting during this current wave of building and investment that is reshaping telecommunications networks globally as voice and data delivery systems converge into a global fiber, and Ethernet based TCP/IP Internet. Therefore although we interviewed Equinix a bit more than a year ago, we decided to return for an update.

**Editor's Note:** Jay Adelson is currently Chief Technical Officer of Equinix. He was the operations manager at the Palo Alto Internet Exchange (PAIX) where he helped to design the PAIX. In June of 1998, he left what had by then become Compaq Computer Corporation to start Equinix with Al Avery, co-founder and CEO of Equinix. We interviewed Adelson on the start up of Equinix in our December 1999 issue. As a Member of Research Staff at Equinix, Lane Patterson runs the Equinix Sandbox research testbed. Prior to joining Equinix, Patterson played a key role in the operations of MAE-East for MFS, and served as Director of Network Management Systems for GlobalCrossing's IPnetwork. We interviewed Lane and Jay on October 20, 2000

**COOK Report:** Having discussed the Equinix business model with you a year ago Jay I would like to focus on a series of technical issues that will soon be involving Internet exchanges and the role that they play in the continued scaling and growth of the Internet. And of course one of a key issues is peering. Exchange points serve as a place for ISPs to interconnect with other ISPs of similar size in traffic and therefore to keep as much traffic as possible at a local level rather than paying for it to go across transit backbones. How are peering issues playing out these days? One gets the impression that they're not quite as controversial as they were year or two ago.

## Peering Issues

**Adelson:** One of the issues facing Tier One and Tier Two Internet service providers is deciding whether to interconnect privately or at an exchange point. Another issue is that

the dividing line between the Tier One and Tier Two is often not very sharp. Based on the number of network routes that they announce, it is relatively easy for me to write up a list of five to ten Tier One ISPs that most people would agree belong in that ranking. The ones most commonly mentioned in the Tier One groups are UUNET, Sprint, Cable and Wireless, Genuity, AT&T, Level 3, Qwest, and Verio. These are ISPs that seemed to have done very well establishing themselves at a high-level of interconnections. Nevertheless, from an engineering perspective, a network is a network. Consequently while there may be Tier Two ISPs that have larger infrastructure than a Tier One, they may not feel like they're part of the Tier One club.

Now the Tier Two ISPs are very highly motivated to peer with each other in order to reduce what they have to pay Tier Ones for transit. Also many of the Tier Twos are motivated to struggle to achieve peering with Tier Ones. Many succeed although sometimes just for historical reasons which often involves their purchase of networks with a peering relationship with a Tier One.

**COOK Report:** Wouldn't many Tier Two networks get peering with one or two or three or even four Tier Ones but wind up having to pay transit into one or two others?

**Adelson:** Yes. I think that from a network architecture point of view unless you peer with everyone, you have to buy transit from at least one network. This is true because, when you get traffic that it doesn't terminate in one of the networks that you have a peering relationship with, you have to find someone to pay to deliver it wherever it needs to go. You must have someone to whom you can default when none of your other peers can deliver traffic. To keep their transit costs low Tier Two networks are motivated to peer as much as possible. Tier Ones peer with each other certainly because their networks have some kind equitable value to each other from the network engineer's perspective and sometimes also with the Tier Twos.

**COOK Report:** Tell me a bit about Bill Norton who from what I hear is really an outstanding expert on the whole area of peering and what he's doing with you. Evidently peering is not quite as contentious as it was a year ago and there has even been some progress made in adjustments of content verses bandwidth.

**Adelson:** There are actually two people and our company who have been very focused on the peering question. One is Bill Norton who has written several white papers on the subject and who has been very focused on helping to educate service providers and the industry on what drives peering decisions.

Sean Donelan who works for our design group is also an Equinix employee. Coming from a number of years of employment with DRA Associates, he has also been very involved from the underdog perspective in understanding the motivations of Tier Two service providers.

Bill Norton has created a lot of models which systematically prove the economic justification for peering at an aggregation point such as an IBX exchange as opposed to private peering between two players with a pair of leased circuits. This is not just because of the quality issue or the expandability issue but also because after you have sufficient fiber in one place, it will attract more and by increasing available options will pay for itself.

**COOK Report:** Tell me a little bit more about Sean Donelan's role in Equinix.

**Patterson:** While Sean's expertise in the dynamics of peering is important to us, one other area in which he is also a specialist in is data centers and how to build them. He has important strategic and tactical responsibilities in a data center design, in networking, and in customer relations.

**COOK Report:** If you're going to open an exchange point, before you do so, don't you almost have to get a critical mass of customers in order to attract other customers?

## Requirements for Being a Connectivity Player

**Adelson:** We have a general rule of form that before we open a new IBX we must have four or five at a minimum fiber providers who connect to our facility. As far as ISPs are concerned, we want to be sure that we have a number of the Tier Ones and as many Tier Twos who are interested to come play. It is indeed a chicken-and-egg kind of issue. We've set up some strategic relationships with some of the key players in various categories of Internet infrastructure. For example content distribution networking, storage, and data CLECs and ISP and carriers whom we can count on to be installed and operating as customers on the day we open the facility. They are paying customers who have agreed that there is value for them as well as for us to come to multiple sites and have signed multiple site contracts with us.

**COOK Report:** What is the average number of fiber providers in each exchange that you now have open?

**Adelson:** I believe it is an average of 7 per exchange.

**COOK Report:** If one is looking at what one must do to be a provider of Internet services

at the infrastructure level today, you must have access to fiber, you must do something about network interconnection, and you must have some kind of coherent business model. How does that stack up as a look at the basic raw operational criteria?

**Adelson:** To answer your question at a high level, I think that today in order to be in Internet connectivity player and there are different business models which can be effective. Furthermore not all of them require that you peer. Case in point being InterNap which buys transit from a number of players and resells it through a kind of an enhanced architecture. But there are various Tier Three, Tier Two, and Tier One players that seek peering as an essential component of carrying on their business. *The reason why exchange points have come back into the forefront of the Internet dynamics is at the speed at which people need to interconnect with each other.* The time as well as the cost that it takes to lease either dark strands or lit strands one to one private interconnect and peering basis of two years ago is now prohibitive. Doing so in a "meet me room" or a central location such as an IBX is really the only solution for a number of these players.

The difficulty is that in the time that it takes to provision a traditional TDM based service from an ILEC - OC -3s and OC-12s and even DS3s take forever. Another consideration, especially if you are a Tier Two or Tier Three service provider, is the cost of that access. It is something that you can mitigate significantly if you eliminate local loops.

**COOK Report:** And is part of the cost of that access dependent on the local SONET infrastructure that the ILEC has?

**Adelson:** Indeed a local loop would likely involve expensive SONET TDM circuits.

## Ways to By Pass the ILEC

**COOK Report:** But if you look at the New York, Washington, and San Jose metropolitan areas to what extent are your ISP customers able to locate at an IBX and bypass the ILECs as they do so?

**Adelson:** Let me explain it this way. Ultimately one thing remains true. Even at our IBXs you have to get in and out the door by means of a fiber provider. Now such a provider could be an ILEC or a dark fiber provider or long-haul carrier or it could be a metro area networking company or MAN like Telseon or Sigma. You have a number of means to get in and out of the exchange. I would suggest that the key consideration is how do you pay for that access in and out of the door of the exchange.

The thought process behind gravitating from a time division multiplexing (TDM) model to leased fiber model — that is from an IXC or leased-circuit to fiber which the user controls is that, if you possess your own glass, you become able to operate with high efficiency. In other words you do not have to pay on a per packet basis. And if, you can afford the initial IRU, getting one gives you the most bandwidth for your money.

If you are going to light your own glass and don't need more than a gigabit per second, gigabit Ethernet is the most cost-effective way to go. You can find many other services. Companies will light strands of fiber for you. They will sell you wavelengths, or it will sell you various other encapsulated protocols in addition to traditional TDM circuits. All these options would come at a lower cost per megabyte than a TDM connection to an exchange that you would get to by means of a circuit from an incumbent local exchange carrier.

**COOK Report:** Suppose I'm someone like Panix in New York City and I have a rack of equipment at 60 Hudson Street. Now suppose I would also like to get a presence at your IBX in Newark NJ. I guess I could do that by jumping on a piece of fiber from Metromedia and going underneath the Hudson River into Newark. But what would be the case in any of the four additional cities you are in and a city like Chicago which you are not yet in? Is it generally possible for ISPs in these cities to come into an IBX exchange or into another exchange and be able to avoid time-division multiplexed SONET-based ILEC fiber?

**Patterson:** It is rather easy to do this. Especially in the major markets, metro area networking opportunities are available today. I've also seen some evidence to suggest that even the traditional ILECs are moving toward marketing some services to compete with the new metro area fiber providers. I can say this because I've seen at least one, if not two Tier One ISPs, leasing OC-192 from incumbent local exchange carriers in order to get into our exchanges at high bandwidth rates more quickly than otherwise possible. The price wasn't what you once would have paid the ILEC, but it still wasn't cheap. This is because the issue was get me in there as quickly as possible with as much bandwidth as possible and, in cases like that, cost sometimes comes in second.

**COOK Report:** Presumably if at an ILEC I'm in charge of certain assets like fiber and add some extra strands that are not gaining income, I'm going to want to explore possibilities for their use and make a deal.

**Patterson:** Carriers will generally get the most income from their fiber by selling lots of aggregated end services such as TDM,

voice, and corporate IP access. However if carriers have extra fiber above and beyond their projected needs, they certainly like to sell it as a long term IRU to get a 20 year chunk of revenue up front. As more players enter the dark fiber market, lease terms are getting shorter.

**COOK Report:** Nayel Shaffei called me some months back when he was forming Enkido and told me that he had acquired no less than 300 route miles of dark fiber from Bell Atlantic in the New York metropolitan area.

**Patterson:** I am hearing stories like that all the time. I am hearing a lot of stories about traditional telcos coming to realize that their managed-SONET-services VPN market may be disappearing. Consequently they begin to sell strands of fiber on their networks or, hopefully, if they're smart, wavelengths instead. If they sell strands, they increasingly run the risk of finding out how much traffic their competitor will be able to put on one of those strands.

## Mechanics of Interconnection

**COOK Report:** Would the next point to cover then be the physical environment in an IBX exchange? I understand that you come in there on fiber, you find a switch and a local switching fabric along with the option of some direct interconnects to other ISPs, and content providers, storage providers and the like. Would you summarize how all this works?

**Adelson:** When you bring services into an IBX, you generally demarc those services in a cage or a cabinet. From that point you can interconnect with other players through direct wire such as a piece fiber that we install and lease to you on a monthly basis. You then connect it to your equipment.

**COOK Report:** Suppose UUNET is in the next cage over from mine and I want to do a direct interconnect. Can you tell me what it's going to cost?

**Adelson:** It will be the same no matter whether it's copper or fiber — about \$500 a month.

**COOK Report:** Presumably that is a lot less expensive than what a customer would be charged at one of the MAEs?

**Adelson:** The MAE's business model is different in that they really do not encourage private interconnection as much as they encourage using the central switching fabric. The idea at Equinix is to encourage people to interconnect by any means necessary.

**COOK Report:** But AboveNet at their exchanges also encourages direct interconnection. Correct?

**Adelson:** Yes. I think that many of these other businesses that have telephone company associations or bandwidth company associations motivated to have you use their network facilities to enter and exit from their exchanges. In our business model you bring in your wire, which could even be copper, and once inside you may connect to other customers or to our central switching fabric. There is no prescribed or favored regimen for you to follow. For example we offer gigabit Ethernet today as a central switching fabric. You can buy a 10/100 port on the gigabit fabric or a Gigabit port at significantly lower price than the competition's. What we're doing now is creating a critical mass on a gigabit Ethernet network inside our facilities.

**COOK Report:** Presumably in about a year with the arrival of more ten-gigabit Ethernet products you will move that fabric up to ten-gigabit speed?

**Patterson:** Oh absolutely. We are always researching various new technologies such as, for example, optical cross connect technology. We are looking at what ever would be appropriate for the next generation of transport. Today the majority of Internet service providers are asking us for a hundred megabit port or a gigabit port on a gigabit Ethernet switch. What you are paying for with our charges is not for the delivery of IP traffic itself but for access to an open market medium where you can, if you want, buy your bits from another Equinix customer.

We put a gigabit Ethernet switch in the middle of each exchange and then we put 10/100 switches with gigabit uplinks on them at various zones. the purpose of these zones is to make sure any customer can reach an ethernet switch within the standard 100m distance, if they are running on copper. If you run on fiber, you aren't distance-bound, and can always connect to the central core switch.

Then if you buy a 10/100 connection we run a cat 5 copper connection from your equipment into that 10/100 switch. On the other hand if you want a gigabit connection, we run a gigabit line directly into the core fabric. Generally we get 10/100 switches in populated with a set number of interfaces and when we run out of those, we simply add on more 10/100 switches.

**COOK Report:** Does anyone in any of your exchanges ask for any kind of a fabric other than Ethernet at this point?

**Patterson:** Yes. We have some demand of for SONET based interconnection and we

plan on introducing those resources for a very high speed interconnects on the order OC-192 or ten gigabits per second. We also have requests for ATM and for high-speed frame. We think it will not be long before we see some demand for the interconnection of optical wavelengths that we plan to be ready to support when the demand becomes significant.

With both SONET and gigabit Ethernet services, the approach is to run a larger core switch, and distribute edge switches around the facility to handle the copper distance limitations. Folks who want high-speed fiber connections go to the central core switch; slower copper connections get distributed out to the edge switches in each zone

**COOK Report:** If you had two big players on fiber who wanted to do an OC-192 cross connect and wanted to make no use of your other fabric, which you agree to that?

**Adelson:** Certainly. We'd love that. Two guys in the same room. OC-192? Or OC 48? No problem. The only charge to them is \$500 per month. Where the central switching fabric comes into play is when you are doing one to many.

**COOK Report:** To pull some figures out of the air, if I am an ISP and I'm considering locating at your exchange, I would review your customer list. Out of ten other ISPs that are connected to that switching fabric I might decide that I could very likely get peering with six of them while having essentially to pay for only the cost of a single hook up. Correct? And can you tell me about how many ISPs are hooked up at your currently open exchanges?

**Adelson:** If you took the total number of ISPs at all our open exchanges whether for peering or web hosting or other purposes, you would get an average of more than two dozen per exchange. You also have companies like UUNET and Cable and Wireless selling transit at our exchanges.

**COOK Report:** Would you describe your relationship with UUNET? Is it just a typical example of a your relationship with the Tier One players in general, or is there something special that stands out about it?

**Adelson:** Our relationship with UUNET is just an example of several relationships we have with large ISPs that are divisions of major carriers and who participate in our facilities. In the case of WorldCom, and AT&T we really wanted those two guys to seed our new facilities. Consequently we undertook strategic agreements with them as well as with Level 3. When we opened the doors on a new IBX, we decided that it is very important to us that companies like UUNET, AT&T, and Level 3 already be in

that each new facility and operational there. The reason for this is that we learned in with the experience of our first facility that, for every month it takes these carriers to become active in an IBX, means that that it is just one of month longer before it becomes a really useful exchange point. We've established strategic agreements with these carriers whereby we have eliminated the risk that we will open an exchange with out them.

## Needs of ISP Customers

Now the process of WorldCom populating an IBX is a two-staged affair. WorldCom owns MFS and UUNET. MFS is the division of WorldCom that will install OC-192s and OC 48s worth of time division multiplexed IP capacity in our exchanges. UUNET operates completely independently from MFS. UUNET will then come and install routers on the MFS circuits. The MFS UUNET relationship is very similar to the MFN AboveNet relationship. Or, before they spun it out, similar to what Genuity (BBN) was to GTE. You often have an IP division of a carrier made responsible for its own infrastructure.

We want all our new sites to have both fiber and IP. What we did then was to sign multi-site contracts for the first couple of providers of each. Once we were sure that our strategic partners would be in each new facility, we found that the rest of our customers came rather quickly because the critical mass necessary to make the exchange worth connecting to was already there.

**COOK Report:** If you're looking then for ISPs to populate your new exchanges, most of the five to ten globally pervasive Tier One ISPs are going to be well aware of what you offer. A major segment of your market presumably is to be found in the 70 to 100 Tier Two ISPs with significant but not global backbones. How do you approach this market? What determines their pattern of existing interconnection? Are any of them likely to move out of exchange points where they're already located and into an IBX?

**Adelson:** We find that if you are at an aggregation point of any kind, you will very likely stay there. You will also, we hope, look at connecting at an IBX to enhance your already existing connectivity. These Tier Two players in you're asking about generally use at least OC -"X" sized circuits in their connections and sometimes leased ATM across pre-existing ATM meshes or frame relay meshes. Although they do have cross-country IP backbones, their traffic is not in any way limited. It would be if false for me to say that Tier Two ISPs as a whole are building networks of smaller overall capacity than Tier Ones. The Tier Twos are pretty big and there certainly open to peering with each

other because each time they do so, that is less money that they have to pay for transit costs to a Tier One.

I would say the Tier Two providers have a great incentive for increasing numbers of Internet exchanges to come online and to be ready to facilitate their interconnection when and wherever needed. As bandwidth usage increases, I expect to see many of these Tier Two providers graduating from an OC circuit range, to the use of an individual wavelength, and eventually to the use of the entire strand.

Although there is merger and acquisition activity going on within the industry, with the emergence of new players, the industry is also growing continually. What we believe is important to potential ISP customers is that our IBX infrastructure allows us to basically eliminate the cost issues of setting up peering with as many other Tier Twos as possible at an IBX exchange. And their bandwidth infrastructures are not necessarily smaller than those of the Tier Ones in terms of the number circuits that they use. They also run into the same time-to-provision issues and availability of resources issues that the Tier Ones do. The cost structures of an IBX that work well for the Tier Ones also work out well for them.

## Metro Area Fiber Providers

**COOK Report:** From your perspective what does the build out of transcontinental and the metropolitan area bandwidth infrastructure look like? In other words from your perspective how do you evaluate companies like Metromedia, Yipes, Cogent, Enkido, Sigma, Telseon and others? What does the glass infrastructure business look like from your point of view?

**Adelson:** There are not too many new guys out there playing with transcontinental glass or even metropolitan glass at the moment. Aerie is doing a major transcontinental, pipeline-based build out with cables containing in excess of 400 strands. PF.net is doing a significant new and largely pipeline based build out primarily in the southern half of the country.

If you are still going to build out, the trick to your success is establishing rights of way, getting the permit authorities to approve putting your class in the ground, and having the requisite amounts of capital needed to buy the glass and equipment. Build-outs are one business model but they aren't the only one. The remaining players in the glass infrastructure industry fall into what you might call a "lease or swap" business model. This often takes the form of swaps based on the premise that if I own glass infrastructure in

one area and you own it in another area, why not swap some strands from each other's network and save each other the cost of having to build out new strands into the other's territory.

Moving on from layer one to layer two, you will have a situation where the people who are building these networks will often light strands with various technologies, and in addition, lease it dark to various other people who do the same. You also have cases like Williams and others who are leasing wavelengths as well. As a result there are various ways in which I can use that glass asset. The use of my glass is important to keep an eye on because, over time, the cost of bandwidth is dropping as it becomes more and more commoditized.

The owners of glass realize that they must concentrate on services that will continue to bring them revenues after they have reached an era of fully commoditized bandwidth. Let me make a couple of comments on the nature of fiber builds themselves — especially from the point of view of your recent interview Bill St Arnaud. A lot of future opportunities depend on people having ready access to their own dark fiber. Certainly to be able to enjoy the benefits of DWDM directly, you need your own dark fiber.

**COOK Report:** And in metropolitan areas how do you go about doing that?

**Patterson:** For one thing we need to be grateful to Bill for his efforts to track this on behalf of the community. Certainly you are seeing more involvement on behalf of city governments after their streets have been dug up by one company after another after another. City governments can start to coordinate fiber builds on a regional basis or, in some cases, they can start to lay the fiber themselves. I think it is exciting to see this kind of involvement because it drives the cost down and pushes equal access to that fiber out to everyone.

**Adelson:** You also need to realize that any business can get its hands on dark fiber by, for example, calling Metromedia Fiber Network and saying that it needs access to a ring from point A to point B. It may take a few months to get it to installed but you are as good as gold since you get a recurring price. I think that once we did get a quote on fiber for the ten mile distance from our IBX facility in Ashburn to Tyson's Corner Virginia was actually \$10,000 a month for a ring. That is really not very expensive. It's becoming very affordable for almost anyone to get into the glass market. But there is only so much glass in the ground, so if you are someone who owns that glass, you want to push your customers if possible into using wavelengths rather than acquiring strands.

Don't forget that the bill is for a ring. That is two pairs of glass each taking two different paths to a destination. You used to be talking about obscene amounts of money for rings. Furthermore the \$10,000 a month price for the ring was quoted at about a year ago. As a result the price may have come down even more. Now with DWDM the efficiencies we can get over those strands are significant. You can see how, although it is becoming accessible to everyone, fiber is still a limited resource from which you are driven to get maximum efficiency by lighting it instead of selling it dark.

## Metro Area Business Models

**COOK Report:** What can you tell me about what you are seeing of the business models of the Yipes, Sigma, Telseon and everyone else in metropolitan areas? Are they best thought of as intermediaries between the Fortune 1000 and everyone else in these areas? Do they have the goal of helping the Fortune 1000 move their traffic from SONET-based local exchange carriers on to corporate run gigabit Ethernet IP networks running IP over fiber?

**Patterson:** Let me answer you this way. I believe that there are a number of different metro area of networking business models. Some involve leasing fiber to office buildings and owning that fiber. And then running various services such as gigabit Ethernet or TDM type services over that fiber rather than just one specific service.

**COOK Report:** Does Nextlink, which, after its acquisition of Concentric, about a month ago became XO.Communications, fit into this model?

**Adelson:** Yes I think they do. Traditionally they have been associated with TDM based services while they have recently acquired a very full telecommunications arsenal. But there are also metro area networking companies that focus a little bit more closely on how they demarc. They will put some kind of smart piece of equipment on the customer premises and will concentrate on using VLANs over Ethernet as a delivery mechanism. They are in effect leveraging the low cost of gigabit Ethernet over those strands and using VLANs as a way to carve the bandwidth up.

They will often throw in enhancements like "high-security" as a way to make their offerings more attractive. Other companies will provision you with just raw Ethernet. Then there are companies specializing more in the traditional TDM world and these companies claim to fame is the speed and low cost at which they can provision those same traditional circuits.

**COOK Report:** And where in this spectrum would Yipes, Sigma and Telseon lie with their business models?

**Adelson:** They are very heavily focused on gigabit Ethernet. These companies start out by leasing fiber and, as they progress, they look at other ways to acquire fiber and extend their networks. One of the main similarities between them is Ethernet hand off. Another similarity is a kind of just-in-time provisioning as well as customer controlled provisioning were you can log into a web site and with a mouse click, up your fast Ethernet port speed from say 45 megs to 50 megs. Because you don't have to wait weeks for changing your terms of service, this kind of provisioning is a big advantage over the traditional way of doing that is characterized by slow service and incremental locksteps from T1 to T3 to OC-3 when increases in bandwidth are needed. In general all of the metro players are solving - one way or another - the problems of faster provisioning and more flexible handoff speeds.

Some of the differences are how they address questions of whether or not they will be an ISP, or offer only layer two services. And in the latter cases various ISPs can do wholesale through them very similar to the North Point or Covad DSL model. I believe that Yipes is looking to provide layer three handoffs whereas the others are focusing on providing layer two services such as VPNs between corporate buildings. Now Cogent has a layer three model that looks as though it will involve having gigabits worth of traffic on its network. Since Cogent is promising people to get those gigabits sent anywhere on the Internet, they have some interesting issues in what they have to do in order to scale their peering and transit networks.

**COOK Report:** In other words one will have to wonder what kind of peering deals the Cogent CEO can strike with UUNET when he brings his hundred megabits per second bandwidth into an exchange point?

**Patterson:** That's correct.

## Further Evolution of Peering

**Adelson:** Peering used to be a very hot political issue. From a technology point it is just as hot as it has ever been. When you are building out your infrastructure you simply must establish BGP relationships with your peers and transit relationships with others. What has changed somewhat now is that a lot of business models are assuming from the very beginning that they will have to buy a certain amount of transit instead of having to depend on establishing hundreds of peers in order to get by without doing it.

You certainly have to begin by buying a certain amount of transit. I remember few years ago when Qwest and Level 3 were ramping up, they always had the question: "can we start as a peer?"

**COOK Report:** And it was very difficult for them to understand why they could not start out as a peer?

**Adelson:** Yes, but the transition to understanding happened fairly quickly. And I think that one of the things that spurred the transition was that someone with \$6 billion in their pocket had as much difficulty in establishing peering relationships as someone with only \$100,000.

**COOK Report:** Well the Level 3 guy with the 6 billion had the flexibility of offering the more established but fiber poor Tier Ones like Sprint and UUNET in return for peering things like "cold-potato" routing, instead of traditional hot-potato. In cold-potato routing, a new player will offer to carry traffic as far as possible to the destination, before handing it off to the receiving ISP. They can do this because they have capacity to spare in both directions. Am I correct?

**Adelson:** Absolutely. Level 3 could enter into a settlement where at some future point if the routing and bandwidth changed direction so did the dollar's. These are examples of the interesting settlement types of relationships that were established. The only problem is that generally no one is willing to talk about them.

**COOK Report:** Well given the contacts that Bill Norton has is he about as good as anyone in getting hard data about what is going on?

**Adelson:** Let me make one thing very clear. Bill Norton, as an individual, has a profound interest in helping the community understand the issues involved in peering and helping players come together in order to peer. As an individual he is been involved in these issues since long before Equinix existed and he will probably continue to be involved in these issues outside of Equinix. As a member of the Equinix team we look to Bill to help to present to the community, in a context that is a little bit larger than this other world, issues associated with cost and other areas of research that he has focused on and written white papers about. In other words, things that assist us with our business model. We are not in any way using Bill Norton as a means of pulling in customers or peers into the IBX exchanges.

**COOK Report:** But there certainly is generic knowledge about peering that Bill has that has to be awfully useful.

**Patterson:** That is true. Bill is able to give

us a lot of generic data about how people want to peer these days. What technology they're using. And what people pay to their upstream providers to get from point A to point B. That information has been extremely viable to us in planning our locations and an understanding marketing strategy. However when I say this, I think it is very important point out that Bill has been extremely respectful of the privacy of individuals involved in these arrangements.

**COOK Report:** Enkido is supposedly providing OC-768 (40 gigabits). Are you aware of anyone else doing anything at this speed?

**Patterson:** It is starting to show up and people are testing it.

**Adelson:** I would say that in terms of traditional Internet service provider infrastructures I haven't seen any lit — yet. Remember that this is a SONET service and, in comparison to gigabit Ethernet, very expensive. Since there are as yet no routers that can handle it, it has to be a point to point service. When you are looking at is doing SONET muxing an OC-768 in order to chop it into manageable sizes.

## Who Needs Interconnection

**COOK Report:** Look at the need of these various fiber metro networking companies about which we been talking, to connect with the networks of Fortune 1000 companies in several dozen metropolitan markets nationwide and deliver them to the internet while bypassing the ILECs. They must get this traffic back out to the Internet. This means that they must go to places like 60 Hudson Street or into some of the existing exchanges and, from your point of view, probably into some of your exchanges. Right?

**Adelson:** Increasingly so.

**COOK Report:** Then if you examined growth in the metropolitan areas where you are opening exchanges, where is it coming from? In looking at the growth, is it fair to segment these people as roughly half the market and public ISPs as the other half?

**Adelson:** I am not sure that would want to characterize it as broken down that way. There are some extremely interesting high-speed long haul plays coming about on the part of some next generation carriers. Nevertheless, at the end of the day, one of the services that a large carrier provides is reach. If you need bandwidth to Greece or to Romania you will not be able to get that from these new players. It takes an extremely long time to build a broad network. These people have very narrow networks: but ones that operate at very high-speed.

**COOK Report:** But for all of these ideas to make sense and to have a truly public Internet, you have to interconnect and these people are your potential customers. And a third area is the content provider.

**Adelson:** Yes indeed a content provider that wants to can control its own network is a good candidate for residence in an IBX.

**COOK Report:** A fourth area would be the application service provider and a fifth area would be remote storage providers.

**Adelson:** Sure. And the remote storage people will be more focused on wavelengths than on traditional time division multiplexing services although we have heard of them encapsulating protocols like Fiber Channel over GigE.

**COOK Report:** Speaking of encapsulation, what are the issues involved when metropolitan area optical fiber players interconnect?

**Adelson:** If they're using gigabit Ethernet as their transport mechanism, they interconnect quite well. The issues that come into play is when they're using DWDM boxes and are trying to talk to other people using DWDM boxes of different manufacturers, the wavelengths may not be in perfect sync with one another. This problem will become mitigated as the standards become more locked down.

**COOK Report:** In the meantime what can be done about this?

## Interconnection - Technical Issues

**Adelson:** We make certain that in our Exchanges we have the necessary expertise to assist customers dealing with these concerns. We help people when they come to deploy understand these issues in advance so, if possible, they can avoid them. In addition there are technologies that we're looking at for future optical interconnection devices, that we hope will bridge the gap. These customers may come in on the inside of one wavelength and then go home out the inside of another.

**Patterson:** There are two sides of your DWDM multiplexor. One is the drop side where you are handing off wavelengths at 1310 or 1550 — one wavelength per port. On the long haul side is where you have multi-colored lambda's on a single fiber. On the long haul side there're still a lot of proprietary issues because everyone encapsulates, in their own proprietary formats, different protocols such as fiber channel, ethernet, or SONET on the edge. Lucent uses something called wave wrappers while someone else might use a different supervisory encapsulation. You could not have a

strand of Glass and have a Ciena box on one side and a Sycamore box on the other and expect the two to be able communicate over the strand.

On the drop side one of the benefits of having essentially a dark fiber exchange model is that people bring in their own fiber and run their own DWDM and then bring it out as a wavelength that they can interchange with others.

**COOK Report:** So precisely how do they do all that?

**Patterson:** Right now in terms of a practical solution, because there are few enough people doing this now, the solution is fairly simple. At an Exchange people simply do dark fiber cross connects between each other. One will say to the other: Hey, I am going to hand you a 1310 SONET based OC-48. However as this type of thing is done more and more, we believe that it will cease to scale with only the use of dark fiber cross connects. When you use a photonic switch that can handle many wavelengths on a single piece of glass, something like a Calient switch with the MEMS system of mirrors, it doesn't care what protocols are coming in and going out. It is just bouncing light from an input port to an output port.

And in fiber switching mode, it is bouncing all the lambdas that it sees on one port (it could be up to a hundred of them) through the mirrors or Mems out to an output port. Consequently it can do fiber switching. It can do lambda switching and there is a whole lot of thought being given to the extent in which it gets involved in the framing of the protocol that is sent out over that particular lambda. This is a somewhat future area but one that is worth while to track. In the meantime, if you want to be protocol agnostic, you can do so. It doesn't matter if it is a fiber channel signal, or a gigabit Ethernet signal, or someone's wave wrapped proprietary signal. It is just a wavelength within an acceptable range.

**COOK Report:** How do you deal with issues of switching versus routing?

**Patterson:** There are a couple of different ways to look at this problem. One: if you are doing a layer one optical exchange it doesn't matter what protocol you were using. It could be MPLS or anything else. Two: if you are running just a gigabit Ethernet fabric, it doesn't matter whether you are passing IP over Ethernet or MPLS over Ethernet.

**COOK Report:** In other words MPLS is something that your customers may decide to run among themselves and may never need to bother you about the fact they're doing so?

**Patterson:** That's right. I think the third point that I should make is that, at this moment, we do not see MPLS appealing to a large percentage of our exchange point populations. But as it moves in that direction, we have been doing testing of LSRs or Label Switching Routers. You can use LSRs in a couple of interesting ways as an intermediate switch at an exchange point. However, there are a lot of issues right now with the use of MPLS between autonomous systems. I guess the short answer for this is that people are still focused on getting MPLS to work well in an in-

tra domain environment.

**COOK Report:** Isn't the real strength of MPLS to be found in the situation where it is used to establish permanent virtual circuits with in a single large network like that of a Sprint or a UUNET?

**Patterson:** Yes. At this point MPLS is best used in doing traffic engineering within your own core. There is an Internet draft for the use of multi protocol BGP extensions to swap MPLS labels between autonomous systems. The motivation for that is to take MPLS VPNs across provider gateways. As an exchange point operator, we must be able to facilitate use of what ever packet switching technologies our customers want.

## What of Wireless and Local Exchange Players

**COOK Report:** Looking at your own experience and at that of any other exchange point operators of which you are aware, what is going on with the incumbent local exchange carriers? Or, for that matter with CLECs? Is it a case of no matter what your business model, if you are shipping bits, you have to jump into the exchange point game and play wholeheartedly?

**Adelson:** Yes. If you are shipping a lot of bits you basically have an economic and technological justification for participating at an exchange. Bear in mind that CLECs can merge. That some are doing so and being successful. The few that exist and are becoming a larger and as such they are also becoming more and more needy of high bandwidth interconnection. We do have them as customers today. We're watching how their value-added service plans are starting to morph. It is a market that is fascinating to watch because they are giving a tremendous value add to the consumer, but the way that the debt markets work controls their ability to succeed because they are very dependent on capital intensive business plans.

**COOK Report:** Looking at some of the other kinds of infrastructure in connectivity players, I know that you have both wireless and satellite dishes on the roof of your exchanges.

**Adelson:** For sure providers of various types are making their way into the IBXs. You have satellite providers distributing content. You have people using wireless line of sight technology to replace circuit technology. You also have people using satellite technology to replace long haul circuits. Just as we cater to fiber coming in and out of our doors, we make certain that we cater to wireless as well. While I don't see wireless replacing core backbone bandwidth any time soon, it certainly is playing a significant role in content distribution.

**COOK Report:** Wireless access providers certainly must interconnect with the fiber infrastructure. Is an exchange point a good place to do that?

**Patterson:** If you look at the GSM or PCS, you find that they have a number base stations around a given city. They will have wireless segregation at a very low level. Consequently they may have 20 sites in a Metropolitan area were they

aggregate wireless into land lines. Once they get it on fiber, they are a carrier just like everyone else in the sense that they have aggregated their traffic onto fiber and are using traditional transport technologies. At that point of course they could route their fiber into a IBX for their interconnection needs.

**Adelson:** I should point out that the wireless world and particularly the cell phone world is moving toward the world of IP and especially toward IPv6. There has been a lot of focus on preparing the critical infrastructure to support this expansion in our exchanges. We would expect that the wireless companies would be aggregating their traffic into very large fiber pipes brought into our exchanges. Once they were there we would serve them by providing a hub for their operations. What we think is really interesting is that we believe that such a hub may well be an IPv6 hub or even a WAP gateway because we do have several customers in the WAP business.

**COOK Report:** Well will the antennas on your roofs be used for outgoing wireless content distribution? Such as for example LMDS?

**Adelson:** People can use our exchanges in all sorts of ways and this is certainly one way about which we have talked to customers.

**COOK Report:** Moving all the way from individual exchanges at this point what do you see as critical issues in scaling large backbones?

**Patterson:** We find so many players with different backbone architecture-religions literally coming into our doors that I don't think it would be safe to comment on seeing any particular trend. But I can say that one of our goals is to be architecturally agnostic and yet be experts at the various architectures in place. This puts us in a good position to identify new interconnection scenarios between network borders.

## An Assessment of OBG

**COOK Report:** Do you have an opinion on Bill St. Arnaud's proposed Optical Border Gate way Protocol?

**Patterson:** I think it is a fantastic concept for a test bed. I also see where his motivations are coming from. He is looking at an environment that they've already reached in Ottawa. There they have significant amounts of their own dark fiber available to them at the edges of the network. Their big problem is one of transit costs. And if they have a lot of large flows going to one fixed point (for example if they want to get across the world to a another large research network like the Star Tap in Chicago) why then pay transit costs for your traffic to get there? Paying for transit gives you the right to go anywhere in the world — that is anywhere in the entire universe of IP routing tables. But if all you really want is to get the bulk of your traffic to is just one, two or three destinations, getting it there by transit is not the most cost effective way.

This is where Bill is coming from. He is also looking at the direction of technology trends. He is thinking: what happens when we can do a thousand lambdas on a piece of dark fiber and we

have a lot of dark fiber of the edge and the core? What happens when we have these signaling models out there and available and the cost per lambda is cheap — implying that it tunable lasers are cheap and that the optical infrastructure in general is much less expensive than it is today? When all of this gets to where gigabit Ethernet is today, then we will have some real opportunities to look at some significantly different networking models.

**COOK Report:** How long do you think it will be before these technology developments become mature enough to affect your commercial business model?

**Patterson:** I think it is more than two or three years away. I think it is fantastic that Bill is working on this and that he is doing the code to work it out. Because the burden of proof is really on him to show working code and working implementations. But I do think if you look at the granularity of an OBGp model that you basically need a flow on the order of an entire wavelength speed to a fixed end network somewhere else in the network in order to justify doing OBGp signaling to establish a light path. In other words today if you are not talking about an OC 48 or OC-192 worth of traffic to another destination, it doesn't make sense. Above and beyond just bandwidth needs, Bill is looking at using wavelengths as a means of establishing more peering adjacencies. But at wavelength granularity, there are some tough cost economics at play—my view is there is a going to be a period of a few years where signaled SONET paths at OC-X for many values of X, will be more appropriate for the bandwidth granularity and cost economics that ISPs require.

Now one of the other interesting things that Bill is driving with his OBGp development is the idea of a virtual device. If you are looking at OBGp, you are looking at an example of an optical switch having several instances of virtual routers on it. What this means is that someone who is connecting owns a virtual slice of the switch.

I am seeing vendors pursuing that model today. The point of this is that there is an obvious advantage at a place where people meet to having a big box that everyone's shares the cost of. Now if that the boxes is an Ethernet switch that's run by a third party exchange point, that's great. But if this box is a virtual router that you actually control a slice of, then the idea is even more compelling. Something in between these two extremes may be thought of as something like a route broker, a policy broker, or a bandwidth broker that is run by a third party but with proprietary software while everyone has a secure channel or signaling path into it. For example database connectivity for policy up loads and downloads. The concept is one of a box in the middle acting as a broker among multiple parties. The virtual router concept that Bill's OBGp revolves around, I think it is a really good way to achieve this.

## The Sandbox

**COOK Report:** To conclude then, what is your

idea of the sandbox all about?

**Adelson:** We believe that the high bandwidth interconnections going on at all our exchange points make them uniquely well equipped places to conduct experiments and tryout projects involving future technologies which will facilitate the trading of traffic between the various participants. This could mean anything from software and server technology, to a hardware and routing and switching technology for various projects.

**COOK Report:** And presumably a sandbox as you describe it is something that only a truly neutral business exchange can do - yes?

**Adelson:** Yes. Equinix Sandbox (tm) is trademark

**Patterson:** The whole idea of an Equinix sandbox is that while you really do have a lot of good laboratories that do performance testing of gear at the specification level. But in terms of an operational burning -in environment and testing environment for production level services, we see the sandbox as merely a responsibility to the community that we have to give people a resource for testing their solutions before they go out on their networks. For example some large carriers want to get together and test inter provider quality of service between their networks. You're not going to be able to do this in a laboratory belonging to either of those carriers.

A sandbox is really a cage and a project structured around the equipment in that cage for this laboratory environment. In such a project anyone who is either a customer of ours be it a carrier or ISP, or anyone else — perhaps a member of the Internet research community — someone who is actually not a customer, but someone who wants to come in and collaboratively prove out new infrastructure concepts, such a party can do it. We don't charge for this and we give the participants free and cross connects, cage space and power.

**COOK Report:** But while you don't charge for it, presumably the only people who can play in the sandbox are those people who are already in your exchange customers or are there for some other purpose?

**Adelson:** Only partially true. You must add vendors or people in the research community who want to talk to our customers. For example this would be a good means of testing a new optical cross connect or a new route server technology, a new protocol which you want to test on a router interface, or a new bandwidth broker box. Is conceivable that people might want to use our Sandbox concept to test a new technology to defend against a denial of service attacks. If I wanted to build an overlay network for that, perhaps we would be good test ground for some of the alpha or beta stages of doing that kind of testing.

Finally the sandbox is likely a good place to test content distribution technologies that may not have surfaced yet such as bellwether. Bell weather is thought of as being relevant to a network denial of service attack or a large network

melt down event such as the release of Web data that hundreds of thousands if not millions of people will want. In a situation like this it is possible to arrange an architecture where you have a surrogate server placed out at the border of your network where you interconnect with other networks. And this device could be either pre loaded or loaded on demand with content that is thought to be in demand.

In the event of a denial of service attack or melt down event, you could re direct routing to this device and thus move in the contents away from the core of your network to the border of your networks. In doing so you would take a load off your core routers. We have contracted with Dwayne Wessel who wrote Squid to develop some of this code. We have run some tests of this concept under various scenarios and have been able to prove they can be effective. The details about this are available on the NANOG web site where we did a presentation on bellwether.

## Global Issues

**COOK Report:** Help my readers understand what all this means in a global context.

**Adelson:** I think one of the areas that Bill Norton has been very interested in outside of Equinix is the overall universe of exchange points on a global basis. How they differ in different areas of the world. As well as the existence of different views about exchanges in various countries depending on the level of infrastructure and scaling of the Internet in that country. Bill gets out to Asia a lot as well as Europe. Unfortunately he is not talking with the press about either Equinix or his understanding of the rest of the Exchange industry.

**COOK Report:** That's a shame. But let's conclude then with a reference to your strategy which is very clear from reading you S-1 Form. That June 2000 document leaves no doubt that your strategy is indeed global. Why you have decided that you have to go global in order to do what you want to do?

**Adelson:** The demand is global. The demand is for a neutral party to offer this level of quality of interconnect space - space that is truly neutral and focused on key Internet infrastructure technology at the level of understanding with which we been discussing with you today. We think that this demand is global. And especially so in the emerging markets that are deregulating now where the benefit of having a neutral meeting place popping up and available is very evident to all the parties. Both the international nature of our customers and the international nature of Internet infrastructure drive us globally. For example consider DNS. It is not a domestic thing. It is a global thing. For us to have our infrastructure available all over the world addresses a problem of Internet growth which heretofore has been too much anchored in the United States. We really need to stop thinking of the Internet as a United States and European phenomenon and to move in a direction where we can address scaling issues globally.

# ITU and IETF in Agreement on ENUM Administration

## Letter from ITU To ICANN Blocks .tel gTLD Applications As Competition to ENUM -- Administration Modeled on Neutral Tier 1 Database Holder of Pointers to Records of Provisioned Services

**Editor's Note:** The ENUM process has taken some significant further steps towards implementation since the publication (in mid October) of our December issue that contained the ENUM interview with Richard Shockey of Neustar. In this brief article we highlight the most recent developments. The material that follows is based on conversations with Shockey.

According to Shockey, "communications between the IETF as a professional engineering organization with the ITU Study groups which are trying to solve engineering problems is in fact excellent and on going on any number of fronts. However problems tend to come when you have to deal with the ITU Secretariat which is the organizations political side and has its own view of the world."

"What happened in Berlin with the SG2 meeting with the IETF that concluded on October 26 was in my judgement a raging success. What happened was a recognition by the ITU that its E164 standard is a good thing and that ENUM itself can be modeled along the lines of E164 in a successful manner that does not compromise the security and stability of the Public Switched Telephone Network. And that furthermore that ENUM also can be deployed in a way that deliberately respects the rights and prerogatives of nation states involving telephone numbering."

### IETF and ITU in Agreement on ENUM

On November 1 a document entitled "Liaison to IETF/ISOC on ENUM" and authored by the ITU-T Working Party 1/2, in Berlin, between 19 - 26 October 2000 was published on the IETF Announce list and ENUM working group list. The document conveys the understanding by ITU Study Group 2 of how it agreed to implement E164.arpa in meetings with the IETF in Berlin between October 19<sup>th</sup> and 26<sup>th</sup>. Shockey added: "As of November 10<sup>th</sup> there had been no complaints on either the ENUM or the IETF discuss list. This is usually taken to mean consensus. Therefore it looks as though at this point the path is clear to proceed towards implementation."

For the Liaison document itself readers

should turn to:

Title : Liaison to IETF/ISOC on ENUM  
Author(s) : R. Blane  
Filename : draft-itu-sg2-liason-enum-01.txt  
Pages : 7  
Date : 08-Nov-00

Abstract: Working Party 1/2, of the International Telecommunication Union P Telecommunication Standardization Sector (ITU-T) held a meeting of its collaborators in Berlin Germany 19-26 October 2000. The agenda of the meeting contained several contributions regarding RFC 2916: 'E.164 Number and DNS' from the Internet Engineering Task Force's (IETF) ENUM Working Group - more specifically, the method for administering and maintaining the E.164-based resources in the Domain Name System (DNS) as related to the ENUM protocol. Consequently, in addition to the WP1/2 collaborators, there were several members of the IETF present to assist with the discussion of issues contained in the aforementioned contributions

A URL for this Internet-Draft is: <http://www.ietf.org/internet-drafts/draft-itu-sg2-liason-enum-01.txt>

Shockey continued: "It is important that your readers understand that these developments have been driven by the IETF, ITU and DoC independent of ICANN." From various conversations with knowledgeable sources we have been assured that ICANN has been told what to do and so far has complied. First with the written instructions from John Klensin on behalf of the IAB to Mike Roberts to insert e164.arpa into the ICANN root in September. The second step occurred in October when four different organizations put forward gTLD proposals to ICANN for .tel or variations on the same.

The implementation of the .tels would essentially duplicate the functionality of ENUM. Such implementation would leave registrants for ENUM services with domains that could, according to ICANN's own rules, be hijacked since ICANN owned the name and not the customer. This move clashed head on with the path taken by the IETF and

ITU on ENUM. The ENUM folk did their lobbying.

On November 1, 2000 in a letter <<http://www.icann.org/tlds/correspondence/itu-response-01nov00.htm>> to Mike Roberts, Yoshio Utsumi, the ITU Secretary General told Roberts to back off and not to inaugurate any names that would compete in any way with e164.arpa. After various paragraphs of diplomatic circumlocution the letter concluded: "As I am sure you are aware, the E.164 international public telecommunication numbering plan is a politically significant numbering resource with direct implications of national sovereignty. It is subject to a multitude of national approaches, regulatory provisions, and, in some cases, multilateral treaty provisions. Considering this, governments should be given the opportunity to fully reflect upon how their particular numbering resource responsibilities relate to DNS-based telephony resources."

"In this regard, the ITU is working with the IETF to progress a careful exploration of these complicated issues in the context of its joint work concerning the ENUM protocol. As there are still considerable areas of coordination work needed at this time, until there is an opportunity to further explore the issues within the context of joint work underway and with national governments; it is the view of ITU that it would be premature for ICANN to grant any E.164-related TLD application as this may jeopardize these cooperative activities or prejudice future DNS IP Telephony addressing requirements."

In other words a strong warning not to do anything to harm the usability of the e164 numbering scheme that had been picked for ENUM. The letter was in early another warning shot across ICANN's bow that that the IETF and ITU not ICANN would write the rules for an ENUM implementing TLD. The result is that with the full support of the IETF and ITU ICANN has formally been told by the secretary general of the ITU in strong diplomatic language to back off and not approve any of the four dot tel proposals. Therefore, in view of the IETF's declaration in the spring of 2000 that .arpa was a gTLD to be used by the IETF for infrastructure purposes, it seems reasonable to look at E164.arpa as ordered into the ICANN root by John Klensin and Karen Rose in September as a new TLD that is not under ICANN's

control.

On November 10 ICANN staff released its evaluation of the new TLDs. What the document had to say about the .tel applications showed that ICANN had gotten the ITU message. "Based on application of the August 15 Criteria, the evaluation team believes that none of the four proposals in the telephony-related group should be selected at this time. Each of the four proposals appears not to have adequately addressed requirements for stable, authoritative coordination with the PSTN numbering system, particularly when dynamic-routing considerations are taken into account. (Of the four, Group One, Number.tel, and Pulver/Peek/Marschel are of particular concern in this area.)

In addition, the Group One Registry, Number.tel and Pulver/Peek/Marschel proposals would do little to address unmet needs. Moreover, if a TLD were established in which the service available at URLs was defined by the TLD rather than the prefix, this would likely increase confusion regarding URL naming conventions. Finally, the concerns raised and caution urged by the ITU counsel against establishing a telephony-related TLD until further study and consensus-building within the Internet and telephony technical communities." Editor: the preceding is accessible at <http://www.icann.org/tlds/report/report-iiib3-09nov00.htm>

## ENUM Administration in North America

What follows is our of the current thinking on the proposed administrative structure for ENUM administration. (These paragraphs are based on notes from a conversation with Richard Shockey.)

In each nation the national telecommunications regulators will be called upon to choose the entity that will actually hold the ENUM resource records for a telephone number. The regulators will also have to determine the rights and responsibilities exercised by that holding entity. The public will have to deal with the entity but will have the ability to ask the regulator to step in and provide due process protection for their rights should they be abused by the entity to which the regulators have granted administrative authority.

In the United States it looks as though there will be a two level system. It is expected that there will be a single Tier 1 entity that will delegate responsibility records and will keep a data base identifying who is authorized to do the provisioning of ENUM types of services assigned to every phone number. The single national tier one entity is

there to certify that person x has two things. The first is authority over phone number y. The second is the entity that this person has selected to do the actual service provisioning.

Now tier two ENUM services are to be provided by the owner of the phone number. If that owner is a large entity like a corporation or a university, it may designate its own telecommunications people to provide the services. However not everyone will really want to do this nor will everyone be technically capable of doing it. Therefore it is assumed that companies will spring up to provide these services for a fee to individuals, organizations and small business. The administrative system will be set so that the Tier 1 entity will be the sole trusted provider of data base pointers to the legitimate owners of numbers and the services they have authorized for them.

From a social and economic point of view telephone service is an absolutely critical service for each individual. Indeed ones phone number is becoming a critical global identifier and the government must guarantee that each individual with a phone number will have control both over it and the services provisioned for it. Or in the event the phone number must be changed the guarantee must be for the seamless change of provisioned ENUM services to the new number.

Because ENUM is becoming service control point for Internet telephony in the same way there are service control points in the PSTN, it is intended that in each nation state there will be a single entity identifying the authorized provisioner of the resource records. These resource records are the means of ultimate control. The tier one entity is to be the repository of the services subscribed to by the owner and user of each phone number. The repository has authority to change its records only on receipt of validated instructions from the number's owner or the tier two agent that the owner has authorized to act on its behalf. With ENUM this puts control of ENUM and internet telephony services into the hands of the subscriber and allows intelligence and service logic to reside in edge devices of the internet as the quintessential STUPID Network.

ENUM administration will put control of service provisioning for an individual's phone number in the hands of entities responsible to the interests of that individual though a process of government ensured accountability and due process. The COOK Report concludes that the proposed process is deliberately unlike that of ICANN which assumes ownership of and total control over a domain name. The process developed has been designed to keep control over ENUM

out of ICANN's hands. In other words the administrative model just described has been structured to remove it from the ICANN control and place it into the hands of a regulator responsible to democratically controlled government who will give the owner phone number full control of services attached to such a number. Policy is to be set by the regulator as a part of the political process within each national numbering system. Everything else is administrative. New policy involving the ownership of a number and the control of a resource record cannot be set without the regulator's approval. Other policies may be set by the tier 2 provisioning entities just so long as national policy on ownership and control is not overturned.

In the United States it is expected that the locus of policy will be either the Department of Commerce or the Federal Communications Commission. An early decision of the new administration will be to authorize either agency to issue an RFC for private sector companies to offer bids on how they would design and implement a single Tier One entity for the United States. Responding to such a RFC will take most of 2001. In the meantime setting up and operating test beds for ENUM services is a major priority.

For a similar point of view (minus the ICANN references) readers should turn to

Title	: ENUM Administrative Process in the U.S.A.
Author(s)	: P. Pfautz, J. Yu
Filename	: draft-pfautz-yu-enum-adm-00.txt
Pages	:
Date	: 18-Oct-00

Abstract: This document considers administrative processes for ENUM in the U.S.A. and offers two 'strawman' proposals in the spirit of moving forward the work that must be done to implement a useful ENUM capability. The U.S.A. has implemented number portability; therefore, it is the telephony user that controls the assigned telephone number so long as it maintains the telephony service. While the proposed processes are tailored for the U.S.A. they may be appropriate for use by other countries that implement number portability so that the donor telephony service provider (e.g., the telephony service provider that is assigned a block of telephony numbers before any number porting event happens from that number block) is not relied on for maintaining the delegation information for a telephone number (e.g., the Tier 1 function in the ENUM process).

A URL for this Internet-Draft is: <http://www.ietf.org/internet-drafts/draft-pfautz-yu-enum-adm-00.txt>

# Is-Is Bug Causes UUNET Route Flap

**Editor's Note:** A "rough edge" of routing of the type that Packet designs had been complaining of surfaced at the end of October with the complaint that a corrupt IS-IS packet can and did cause a large Cisco route flap through a major provider network.

On October 29, 2000 **Sean Donelan** wrote to NANOG: Network operators using IS-IS on Cisco should be aware of a problem which can result in their routers reloading unexpectedly when a corrupt IS-IS packet is received. Cisco has fixed the problem in various 12.0 software trains. The problem can cascade through a single provider's network. Because IS-IS is an IGP protocol, it does not propagate between providers. This has already affected a provider. I haven't found out what is originating the bad IS-IS packet, i.e. is this an inter-vender operability issue?

Mentioning Sean Donelan's assertion that "because IS-IS is an IGP protocol, it does not propagate between providers," **Sean Doran** replied: This is not the reason why it will not propagate between separate ASes. The "saving factor" here is that nobody really routes CLNS natively, and therefore, the maximum hop-count of a CLNS datagram is 1.

It would be possible to cascade an IS-IS problem across multiple separate ASes in the unfortunate event that more than one AS treated a single LAN (e.g. an IX) or point-to-point link as an internal one across which IS-IS is run, with the same key. This kind of mutual poisoning between separate ASes happens with some regularity, amusingly often with RIP as the IGP.

An IGP based on a natively routed protocol (including routed CLNS) widens the scope for inter-AS poisoning. This is why it is important to have good authentication in one's IGP. Unfortunately, \*no\* IGPs currently in wide use have any such thing. :-)

For clarity, a separate AS is really short hand for, "a collection of routers participating in a common IGP instantiation"; there are cases where different ASes (in the BGP sense) share a common IGP. Also, "propagating between providers" seems to ignore the fact that there are single providers who have multiple IGP instantiations. P.S.: any chance you can be a bit more concrete about what's happening?

**Sean Donelan** on Oct 30: When I'm concrete, providers complain I'm picking on them, and getting them bad press. But since you asked....

At approximately 7:37am EDT on Friday,

about 258 Cisco 12000's on UUNET's primary backbone reloaded. This appeared to be isolated to routers in ASN 701. It disrupted reachability to about 15% of the world-wide Internet based on data from Matrix measurements. A contributing cause was a bad IS-IS packet which confused certain IOS versions in the 12.0 IOS software train. I haven't heard what the root cause was or what originated the bad IS-IS packet. The Cisco bug id is CSCdr05779. Any provider running the affected IOS version may be vulnerable depending on what the root cause turns out to be.

Although the bad IS-IS packet didn't propagate to other providers, several other providers did report BGP resets and route flaps about the same time.

**Neil McRae:** If a large AS such as AS701 starts flapping I wouldn't be surprised if other ASes start seeing BGP resets and route-flaps. Could be that crud routing information was exchanged when that chaos started [jeez 258 routers I'd hate to have been the on duty NOC guy on that morning :-)]

Interestingly though we still see a lot of routes with bad AS-PATH information people should be setting more stringent configurations on the routes they learn and subsequently pass on to avoid this.

**Roland Dobbins:** I had a bizarre event occur on Thursday night/Friday morning, and this is likely the culprit.

I peer with AS701. At approximately 11:15PM PDT or thereabouts (2:15AM EDT), the 7507 which provides my connectivity to uu.net went belly-up in a very strange way. The BGP session with 701 showed active, with full tables; existing TCP connections stayed up. However, it appeared that all new connections inbound from 701 were being dropped on the floor, and my outbound traffic with them dropped from 40mb/sec down to about 5kb/sec. The same router was also handling a secondary connection to pbi.net; because the BGP stayed active and in a supposedly functional state, traffic didn't get routed in that direction as it should've been.

I had to reload the router to get it to function properly. Very odd. Nothing in the logs, etc. The router just essentially went on strike, and I've no idea why. I don't run IS-IS, needless to say, especially with a foreign AS. This particular 7507 was running an 11.3.x CC-train IOS, and hadn't had any of the ISO/CLNS family of protocols enabled, ever. This is a bit earlier than the timeframe Sean cited, but I don't think it was a coincidence, either.

**Sean Doran** responding to Roland Dobbins description of his bizarre event .

Some of your symptoms are consistent with a badly-broken sloshing IGP, notably the drop in traffic load and large numbers of dying TCPs passing through the afflicted network. This is two sides of the same coin: a destination in your network, learned through (e)BGP is mapped to a next-hop address (typically the interface across which you are talking (e)BGP) and propagated through their network via iBGP. The IGP is used so that each iBGP-talking router knows how to get to each next-hop address. A sloshing IGP will break connectivity between a given router and all the addresses associated with a broken next-hop. A hypothesis: for each afflicted router, the failure of one next-hop-address to be reachable will cause your ENTIRE network to be unreachable by sources relying upon traffic passing through that router. This may mean a sizeable proportion of their customer base simply could not reach you reliably enough to maintain a TCP connection in equilibrium, or at all. Frequent transition to slow-start due to loss/out-of-order-packets \*and\* a reduction in the overall number of TCP "mice", would severely reduce traffic.

An interesting question, however, is why would their iBGP TCP connections appear to remain functional (you aren't losing eBGP routes) in this sort of mess? Did loopback addresses not come and go, but interface addresses did? (That would be interesting to consider in the face of possible aggregation of interface addresses into the IGP). Is there significant partitioning because of, for example, AS confederating, mitigating the problem by removing iBGP's need to know about distant loopback addresses, but not distant next-hop-addresses?

We are lucky to have what could be a very interesting case study in routing scalability trade-offs. What a pity nothing like outage@sprint.net exists any more, where we might find useful information from the victim provider. :-)

**Neil J. McRae:** Indeed. outage@sprint.net still exists BTW - it just doesn't have any other information other than "A router did/will/might be reloaded"

**Dobbins** (in answer to Doran's diagnosis of his bizarre event): No, I'm a single-AS hosting provider, no confederation.

**Sean Doran:** UUNET has confederations; I was doing a public thinking exercise (trying to coopt smart people who still read the NANOG list, too) about the extent to which the insulation of iBGP in a confederated AS

from the disappearance of iBGP peers for an otherwise full-mesh iBGP layout interacts with the exposure of multi-AS-one-IGP confederations to failures which cause the iBGP next-hops to disappear. If it seems a bit esoteric, don't worry, it is...

**Dobbins:** The more I think about it, the more I'm convinced that CEF simply stopped working; all my interfaces were active, and there were no apparent problems with my IGP, which is OSPF.

**Doran:** Right, UUNET was having the prob-

lems; you were just a victim of their internal routing being so broken that they couldn't make packets move to you reliably, even though their routers knew how to get to your network; likewise their routers kept telling you they knew how to get to all sorts of networks which in fact they couldn't reach.

This problem appears consistent with an IGP problem inside UUNET, which is known to use/have-used confederated ASes.

**Dobbins:** I think that major BGP wigginess caused the CEF problem; thanks very much

for you insight, I definitely need to think about it some more.

**Doran:** What makes you think it was a CEF problem?

**Roland Dobbins:** You're right - I'm looking for fault here in my own equipment, whereas a screwed-up IGP on their end could well have caused all the problems I was experiencing. Occam's Razor, of course! Your point about the iBGP confederations is well-taken, by the way.

## ICANN Having No Authority to Create New gTLDs Lacks Legitimacy in US and Is Increasingly Rejected in Europe Dixon Explains How .eu Has Been Kept from Icann Control

**Editor's Note:** As we have contended since last May, ICANN is nothing more than a shell game played with the Department of Commerce to convince the rest of the world that if it behaves the US government will give ICANN independent authority over the root and the DNS system globally. ICANN has accepted 2.2 million dollars in application fees from 47 entities that want to be given global rights to administer new domains. In response the country code administrator for Belize filed suit against ICANN seeking to enjoin it from adding "biz" to the root. On November 13 ICANN posted to its web site a response to the law suit against it stating that it had no authority to add new names to the root after all. Such authority, it admitted belonged to the Department of Commerce.

"ICANN represents that it has no authority to implement new TLDs, and that instead, it merely makes recommendations to the Commerce Department, which retains the ultimate authority to make such decisions. ICANN further represents that before any such recommendation would be made, ICANN would have to successfully negotiate a fairly complex agreement with the selected applicant. " See <http://www.icann.org/tlds/correspondence/esi-v-icann-13nov00.htm>

If ICANN lacks legitimacy and authority in the United States, it is also rejected in Europe. Furthermore many Europeans are expressing more and more strongly the opinion that it simply unacceptable for a California based corporation to even pretend to have the kind of authority that ICANN does over a global resources as important as the internet.

In early November a controversy arose as to whether a web site [vote-auction.com](http://vote-auction.com) had

been removed from the root at the orders of United States officials with the result that the site was no longer reachable from Germany. It lead to a discussion that well identified European concerns about the acceptability of ICANN's existence as an entity subject to California and to US law.

On Fri, 3 Nov 2000, **Tom Vogt** wrote: it seems that core (i.e. the root servers) has deleted the entry for [vote-auction.com](http://vote-auction.com) - while the whois still works and their primary nameserver (in Austria) still resolves, a regular lookup returns with "host unknown".

Rumour has it that Core carved in to demand by most possibly the feds. Here in Europe the sentiment today is that by doing so core has stopped being (if it ever was) an independent and purely technical instance and has entered the realm of politics. for example, no matter whether or not [vote-auction.com](http://vote-auction.com) is or is not illegal in the US, what business has a US court in blocking the site for \*me\* (in Germany) or, for that matter, the rest of the planet?

**Jim Dixon:** Tom Vogt pointed out in a follow-up email that 'CORE' should be replaced with 'InterNIC'. CORE as the registrar actually still had the name listed.

Nevertheless, what has happened here demonstrates a basic flaw at the heart of the domain name system. ICANN and many essential Internet resources remain subject to US jurisdiction. ICANN itself is just a California corporation, so it is subject to the passing whims of the California legislature as well as those of Congress, the executive branches, and various and sundry US state and federal courts.

Some argue that ICANN should itself have authority over all of the Internet domain

name system and the IP address space and in fact things are creeping in this direction. Given the now-crucial role that the Internet plays in the global economy, ICANN's hegemony gives, for example, representatives of small towns in California sitting on the right committee in Sacramento remarkable and truly unique power over the rest of the planet.

To complaints by **Dave Crocker**, As usual: 1. it is vastly easier to criticize the status quo than to propose something superior; and 2. it is vastly easier to propose general ideas than to provide detailed plans; and 3. it is vastly easier to specify a plan than to make it happen.

So what is the point of offering the criticism, absent having done steps 1 & 2, and some of 3, above?

**Dixon** responded: I do believe that this is called begging the question.

Given ICANN's peculiar legal status and vulnerability to law suits, I strongly recommended to the European Commission that steps be taken to ensure that .EU would be delegated as a ccTLD rather than (as proposed) a gTLD under ICANN's new procedures. Fortunately this advice was accepted. That is, we did steps 1, 2, and 3, and in consequence .EU will be largely free from the ICANN mess.

Those involved in actually building the Internet on a day to day basis spend a good deal of time engineering away single points of failure. ICANN is just such a weak point. Having power over the DNS, the Internet address space, and various other essential bits of Internet infrastructure all concentrated in one private company in California — especially this particular private company —

is simply foolish.

Whatever can be done to provide diversity and resilience in the management of the Internet should be done. Keeping .EU clear from ICANN's entanglements was a small but real step in this direction.

On Sat, 4 Nov 2000, **Jim Bell** wrote: But that's not the whole problem, here. ICANN may be, arguably, subject to "those laws," but it isn't clear that those laws (per se) were responsible for the disconnection. Is there a law, somewhere, that said "anybody who we determine appears to be violating the law in America, we 'unaddress' them before they get a trial." That certainly isn't normal procedure: There are probably over a thousand Internet Casinos who are (the thugs would argue) in violation of some American law, yet they are still accessible to us.

**Dixon:** There is a very large world outside of the United States. There is no reason why issues involving .UK, for example, should be subject to the jurisdiction of California courts. Britain is not a colony of the United States, nor is it a California county.

Nor is there any justification for US government control over the allocation of IP address space within Europe. But when you look closely at ICANN, this is what you are getting.

ICANN was supposed to replace IANA. IANA had a narrow technical role that depended upon voluntary cooperation. Having IANA arbitrate decisions about .UK actually worked, because IANA did not claim any ultimate legal authority. It was just obvious to everyone that if they didn't cooperate the Internet would not work.

It may seem odd, but because IANA was gossamer thin, it had real power and legitimacy. ICANN doesn't and shouldn't.

**Bell:** ICANN needs to be taught a very painful lesson: "Even if you feel that you must obey a specific law, you must not do it without initiating a legal process and continuing it through any valid appeal. Given that the election was only a few days away, it is obvious that no such process would be completed before the point becomes moot. You screwed up."

**Dixon:** ICANN is a California corporation subject to state and US laws. It has an obligation to obey those laws. There is or should be no question about this. ICANN is after all a legal fiction, a body whose very existence rests upon the authority of the state of California.

The question is whether the domain name system, the IP address space, and other fundamental Internet infrastructure should be

subject to US and California law. These are global, not local, resources.

**Dave Crocker** wrote: You started by citing fear of legislators in Sacramento.

Now you say the problem "seemed" to be pressure from elsewhere. The problem "seems" to be dancing around.

**Dixon:** The problem is that ICANN, which aspires to rule the global and international Internet, is a private company, a California corporation subject to the whims of the California and US courts, legislators, and executive branches. The success of ICANN as currently structured means that Sacramento and Washington would acquire an unwarranted degree of control over key institutions of the Internet, including, for example, the various European domain name registries.

A second problem is that ICANN is a single point of failure, something that all of us who are involved in the practical engineering of the Internet seek to avoid.

**Crocker:** What is the specific, superior solution that you are proposing?

**Dixon:** That all of us do what we can to resist ICANN's attempt to impose control. That all of us do what we can to build workarounds, to increase diversity in the management of the Internet.

**Bell:** No reason why issues involving .UK, for example, should be subject [to control from the USA]

**Crocker:** No reason other than the absence of a viable, detailed specification for an alternative operation. And consensus support for it.

**Dixon:** The UK \_has\_ an "alternative" operation for the management of its domain name system. Nominet, the .UK registry, is very well run. It's very efficient. Fees are low. It is blessedly free of the endless disputes that characterize the US-based elements of the DNS. Nominet has the consensus of the UK Internet community behind it. It works.

And Europe \_has\_ an "alternative" model for the management of IP address space. RIPE, the European registry, is also well-run and efficient. It's fees are low. It has universal support. It works.

**Crocker:** Really, Jim. Criticisms of the type you continue to offer are entirely wasteful, absent a viable alternative.

**Dixon:** Really, Dave. You just don't get the point: it's ICANN that is not a viable alternative.

ICANN is attempting to force the world's DNS and IP address space registries to accept contracts that give it effective hegemony over the Internet. Fortunately, by and large these Internet institutions are rejecting ICANN's attempt to impose centralized control. What the Internet needs is institutions which foster voluntary cooperation for mutual benefit. We do not need rigid control by a bloated bureaucracy.

On Sat, 4 Nov 2000, **Dave Crocker** wrote: How does another ccTLD in any way "provide diversity" for gTLDs?

**Dixon:** Several hundred million people live in Europe. .EU is likely to become the TLD of choice in this continent. It will attract many who would otherwise register names in .COM/NET/ORG; it's likely that many millions will register names in .EU.

One option was that .EU would be chartered as a new-style ICANN TLD; this would have given ICANN nominal control over what will become a substantial part of the domain name system. Fortunately the decision was to have .EU classified as a ccTLD.

**Crocker:** It had sounded as if you were concerned about that set of domains.

**Dixon:** I do believe that EuroISPA's comment on the US government green paper on the DNS suggested that the best thing to do with .COM, .NET, and .ORG was to push them under .US. In other words, no, I am not much concerned about the gTLDs.

**Crocker:** Your original note and latest response continue to ignore the hard work of providing and pursuing detailed plans to remedy the problems you cite.

**Dixon:** Over the last several years I have spent a great deal of time and done a lot of hard work in lobbying for sensible government policies towards the Internet, both in the UK and in Brussels. In particular, EuroISPA proposed the creation of .EU to the Commission several years ago and has been active ever since in arguing for rational policies in its management. We have tried very hard to avoid the sort of senseless wrangling that has characterized the US-centric DNS wars. Had .EU been classified as an ICANN gTLD, it would have been entangled in those wars. .EU as a European ccTLD is free of ICANN and free of the DNS wars. This is a Good Thing.

This is not to say that there will be no problems in the management of .EU. Doubtless there will be problems; but they will be solved by different people in a different way. That is, the management of the DNS will be somewhat more diverse than it otherwise would have been.

In my opinion, we don't need grand solutions of the type that you seem to be arguing for. What we need are small, practical steps towards greater diversity in the management of the Internet.

This is all becoming a bit repetitious, so with apologies, unless you have something new

to say, this will be my last word on this subject. It was good to see you in Yokohama, Dave.

**Editor's Conclusion:** As it goes into its November Annual Meeting, it is beginning to look as though ICANN disparate to get control over the country code domains to

force them to pay their ICANN dues, is inviting the GAC to step in and force each domain under the control of the government of that name. If the European Commission is as determined as Dixon believes to keep .Eu free of ICANN control, the we would hope to witness the delightful spectacle of European secession from the GAC.

## DNRC Letter Documents ICANN Past Testimony to Show Duplicity Behind So Called Clean Sheet Study of Public Board Members

November 13, 2000

Esther Dyson  
Chairman of the Board  
Internet Corporation for Assigned Names and Numbers  
4676 Admiralty Way, Suite 330  
Marina del Rey, CA 90292

Dear Ms. Dyson:

On behalf of the undersigned public interest organizations, we write to protest the proposed study of the at large membership adopted by resolution at Yokohama and scheduled for implementation at this meeting. The "Clean Sheet" study breaks faith with the United States Congress, before which you testified in August 1999 and promised a democratically elected Board. It breaks faith with the Commerce Department, and those who participated in the ICANN approval process in 1998. Finally, and most importantly, it breaks faith with the Internet community as a whole, to which ICANN repeatedly promised permanent representation *at least* equal to that of the supporting organizations.

### BACKGROUND: ICANN'S SHIFTING COMMITMENT TO THE AT LARGE MEMBERSHIP AND DIRECT ELECTIONS

A brief recitation of the history of the At Large Membership and direct elections is in order here. As demonstrated below, ICANN's enthusiasm for an At Large Membership and direct elections has waned over time. Initially, when it sought approval from the United States government and the Internet community, ICANN professed great enthusiasm for an At Large membership, with the power to elect nine members of the Board through direct elections. Only after its position over the DNS became secure, Congress relaxed its vigilance, and the opposition of NSI was neutralized, did ICANN retreat from its commitment and oppose At Large representation on the Board.

When Commerce first sought to transition DNS management to a publicly accountable "bottoms up" private organization, ICANN pledged to create an open membership struc-

ture to assure public oversight and public input. ICANN promised to create an "At Large" membership which would directly elect 9 members of the Board, as a counterweight to the 9 directors elected by the Supporting Organizations. See *Letter of Esther Dyson, Interim Chair, to J. Beckwith Burr*, November 6, 1998, <http://www.ntia.doc.gov/ntiahome/press/ICANN111098.htm>. That letter contained the following promise from the ICANN Board.

Some remain concerned that the Initial Board could simply amend the bylaws and remove the membership provisions that we have just described above. *We commit that this will not happen.* In addition to our commitment, the U.S. government has publicly stated that it will maintain oversight during the transition period, and *we fully expect that the creation of a membership and the transfer of authority to a fully elected Board will occur before that transition period ends.* (emphasis added)

Based on this assurance, the Commerce Department entered into a cooperative agreement with ICANN. See *November 25, 1998 Cooperative Agreement*, <http://www.ntia.doc.gov/ntiahome/domainname/icann-memorandum.htm>. That agreement explicitly requires that ICANN:

Collaborate on the design, development, and testing of appropriate membership mechanisms that foster accountability to and representation of the global and functional diversity of the Internet and its users, within the structure of private-sector DNS management organization.

ICANN initially appeared eager to make good on its pledge. ICANN formed a special Membership Advisory Committee (MAC) composed of a broad cross-section of the Internet community. ICANN assigned a swift timeline for study, and requested a report for its Berlin meeting in May of 1999. After much hard work and careful study, the MAC completed its task and submitted a comprehensive, balanced report.

In the summer of 1999, ICANN became the subject of criticism for its closed processes, its lack of accountability to the Internet com-

munity as a whole, and its proposed \$1 "domain name tax." Rep. Tom Bliley, Chair of the House Commerce Committee, sent inquiries to ICANN and convened a Congressional oversight hearing. Network Solutions, Inc. refused to recognize ICANN's authority over the DNS, citing ICANN's lack of accountability.

In response to these inquiries and criticisms, ICANN repeatedly promised that establishment of an open membership, direct elections, and the resignations of the initial Board members were its "top priority." See, e.g., *Letter of Esther Dyson to J. Beckwith Burr*, July 19, 1999, found at <http://www.icann.org/correspondence/icann-to-doc-19july99.htm>. In particular, you assured the Department of Commerce (and the Internet community as a whole) on behalf of ICANN that:

Our goal, which I know you share, *is to replace each and every one of the current Board members as soon as possible* (emphasis added).

ICANN made similar pledges to Congress. In sworn testimony before the Congressional oversight hearing on July 22, 1999, you testified:

As to the second wave, *it is ICANN's highest priority to complete the work necessary to implement a workable At-Large membership structure and to conduct elections for the nine At-Large Directors that must be chosen by the membership.* ICANN has been working diligently to accomplish this objective as soon as possible. The Initial Board has received a comprehensive set of recommendations from ICANN's Membership Advisory Committee, and expects to begin the implementation process at its August meeting in Santiago. *ICANN's goal is to replace each and every one of the current Initial Board members as soon as possible.*

*Testimony of Esther Dyson, Chair, ICANN, before the House Commerce Committee, Subcommittee on Oversight and Investigations, July 22, 1999* (emphasis added), found at <http://www.icann.org/dyson-testimony-22july99.htm>.

Congress appeared satisfied by these pledges. In addition, NSI entered into negotiations with ICANN and the Department of Commerce which ultimately culminated in the existing agreement whereby NSI recognizes ICANN's authority and provides it with financial support.

Once this oversight and opposition to ICANN vanished, however, so did ICANN's commitment to the At Large membership and direct representation. At the Santiago meeting in August 1999, the ICANN Board did not implement the recommendations of the MAC or step down in favor of elected representatives. Instead, the initial Board members extended their terms another year, and adopted a resolution to prohibit direct elections of directors by the At Large membership. See *ICANN Board Resolutions*, found at <http://www.icann.org/santiago/santiago-resolutions.htm>.

This announcement prompted outrage within the Internet community as a whole. It directly contradicted ICANN's previous statements, quoted above, regarding the priority ICANN placed on membership, its commitment to public oversight, and its willingness to allow open elections. The cynical nature of this shift was lost on none, yet public interest organizations continued to attempt to engage the Board in discussion over this change of course. ICANN rebuffed these attempts at civil engagement and public dialog, although it gratefully accepted the \$100,000 grant from the Markle Foundation to create an At Large membership. See *ICANN Board Resolutions*, <http://www.icann.org/minutes/prelim-report-4nov99.htm>.

In March 2000, an independent study by Common Cause and the Center for Democracy and Technology demonstrated that ICANN's proposed indirect election was a "risky experiment in democracy that must be dramatically improved for it to confer legitimacy on ICANN." See *ICANN's Global Elections: On the Internet, For the Internet, March 2000*, Found at <http://www.cdt.org/dns/icann/study/>

In response to this report and continued criticism, the ICANN Board agreed to the "Cairo Compromise." Under this proposal, ICANN would finally create the mechanism for general membership, but it would only allow the election of five of the promised nine At Large directors. See *ICANN Board Resolutions*, found at <http://www.icann.org/minutes/prelim-report-10mar00.htm>. In addition, the Board *again* extended the terms of initial Board members, allowing four of them to remain on the Board until October 2001. *Id.*

This proposal reduced the accountability of the Board in two ways. First, it perpetuated

the terms of 4 of the unelected "Initial" or "Interim" Board members, despite the repeated pledges in 1998 and 1999 that these directors would be replaced by directors elected by the At Large. These "Boardsquatters" reduce the number of directors accountable to an electorate.

Second, it allowed any two Supporting Organizations to overwhelm the interests of the At Large membership. The structure ICANN first announced in 1998 allowed the general membership to serve as a check on the narrow representation of the SOs, while permitting the specialized SOs to neutralize the At Large only by complete agreement. Under the new structure, the three Directors from any two supporting organizations can neutralize the votes of the At Large membership.

Despite these flaws, the public interest community applauded the Cairo Compromise as a step in the right direction. At the time, those in Cairo and the broader Internet community fully expected that ICANN would, ultimately, make good on its repeated promises to have an open membership that elects Nine directors.

This hope, however, again proved vain. At its August 2000 meeting in Yokohama, the Board adopted a resolution calling for a "clean sheet" study of the At Large membership. The resolution explicitly contemplates that all aspects of At Large membership, indeed, its very existence, will be subject to re-argument and potential elimination. Recent comments from ICANN President Mike Roberts indicate that, despite the success of the recent elections in electing qualified candidates (although it experienced flaws in implementation), *the Board intends to eliminate the At Large Directors rather than expand them to the full nine promised when ICANN first formed.*

**BY ITS ACTIONS, THE BOARD HAS CALLED ITS OWN LEGITIMACY INTO QUESTION**

By its own actions, the Board has called its own legitimacy into question. ICANN received its current authority over the DNS based on bylaws that promised an open At Large Membership that directly elected nine directors to the Board. In response to charges that the insiders in ICANN would cynically delay, extend their terms, and wait until their positions were secure, were dismissed by ICANN. "We commit that this will not happen," wrote the Initial Board in 1998, and upon which promise it received authority over the DNS.

Two years later, we know the worth of ICANN's commitment, particularly that of the four Boardsquatters who have clung to power in the face of all promises and pressure to the contrary.

The proposed "Clean Sheet" study further vitiates any shred of legitimacy ICANN might hope to obtain from elections. At best, it becomes a means of intimidating those who favor a permanent At Large Membership with the power to elect Directors. "Behave," the Boardsquatters

will say, "or we will eliminate the At Large entirely." More likely, the Boardsquatters and other elements of the Board will use the Clean Sheet study to further reduce the ability of the At Large to serve as an effective means of public oversight and a counter to Board capture by the unelected Boardsquatters.

This state of affairs is intolerable. Unless ICANN takes swift action to redeem itself, it will have perpetrated a fraud upon the Internet community and those who entrusted it with stewardship of the DNS. A house built on a false foundation cannot stand. ICANN's actions undercut its own legitimacy and undermine its ability to move forward as a consensus builder within the Internet community.

**ICANN SHOULD ELIMINATE THE "CLEAN SHEET" STUDY, REMOVE THE BOARDSQUATTERS, AND INCREASE THE NUMBER OF AT LARGE DIRECTORS TO THE PROMISED NINE**

For the sake of its own legitimacy, ICANN must move quickly to make good on its representations. It must reaffirm the role of the At Large, eliminate the Boardsquatters, and allow new At Large Directors accountable to the At Large Membership to fill their seats.

The road to these steps is easy and straightforward. Eliminating the "Clean Sheet" study requires nothing more than a bylaw change, a process with which the ICANN Board is intimately acquainted and has had no shyness in employing in the past. The Boardsquatters have it within their power to resign, which they should do forthwith.<sup>1</sup>

Professor Michael Froomkin of the University of Miami School of Law has ably set forth a plan to replace the Boardsquatters with At Large Representatives, so that ICANN may function with a full compliment of Directors until ICANN can hold new elections. Following the Froomkin Plan, ICANN should permit the five elected At Large Board members to select the four replacements for the Boardsquatters. While Professor Froomkin is right that this is not a perfect solution, it is the most equitable way to solve the problem and allow ICANN to function until it has recovered from its unfortunate attempt to evade its responsibilities.<sup>1</sup> A copy of Professor Froomkin's proposal is available at <http://personal.law.miami.edu/~froomkin/boardsquat2.htm>.

**CONCLUSION**

The actions of ICANN's unelected Directors have undermined ICANN's legitimacy and compromised its ability to function effectively. This naked display of bad faith and ambition cannot help but repulse the Internet community as a whole, and prevent ICANN from claiming the mantle of an consensus rule and "bottoms up" management. If the unelected Board members, and in particular their Boardsquating brethren, have any respect for the organization they have spent two years building, they will cease their assaults on the At Larger Membership and the elected Directors. [Http://www.netpolicy.com/icann111000.html](http://www.netpolicy.com/icann111000.html)

## Executive Summary

### Packet Design pp. 1 - 6

Packet Design is a venture headed by Judy Estrin, Kathie Nichols, and Van Jacobson. It is the fourth company founded by Estrin and her husband Bill Caricco. Buoyed by their first three successes the founders have the ability to run Packet Design with a business model of a perpetual startup. As explained by Judy and Kathie in an interview, while Packet Design does anticipate making money by licensing and spinning off startups of its own, it has a much more unique purpose to its existence. Namely it has been formed to do research into improvements of routing. Judy explains that the frantic growth of the commercial Internet from 1995 onward meant that short term quick fixes were taken to the evolution and scaling of routing and backbones. To her dismay as the path to the convergence of voice and data networks, the PSTN and the Internet is now clearly defined and momentum of convergence is increasing, she sees a possibility that we could wind up with the worst of both worlds in the final architecture.

In her own words: "The convergence of data, voice and video we are seeing today is driven by the dramatic increase in data traffic now being pushed across the PSTN infrastructure. The traffic patterns associated with new data applications are very different from those of phone conversations. Yes, the Internet needs to maintain the manageability of the telephony world -but not at the expense of scalability. MPLS, a string-oriented technology, was developed to solve a point problem: integrating local IP and ATM environments. It works well for that use, but its proponents have positioned it as a panacea for all sorts of other problems. At first glance, MPLS seems like the perfect answer to a converged Internet. But it's really just a quick fix. Because its architecture is based on strings rather than clouds, it has all the disadvantages of strings and, in the long run, it creates more problems than it solves."

Judy answered affirmatively our questions as to whether a part of her message was that she believed that an Internet where the circuit-switched orientation of MPLS (strings) played the major architectural role would be more costly to operate than an Internet that was faithful to its founding connectionless philosophy (clouds).

"The reason that we feel so strongly about this is that we believe that it is the properties of IP which can take us directly to the type of Internet that yields the best scaling characteristics at the most favorable cost from a manageability perspective. Now what the telephony world does very well is to offer us some best practices in the areas of manageability and billing and accountability. We need to map these onto the Internet world. But you don't achieve this by making the Internet infrastructure look like the telephony infrastructure. You need to figure out how to do all those things within the confines of clouds."

While Judy, Kathie and Van believe that they have definite ideas that when implemented will pro-

duce answers to MPLS and improvements in routing, as Judy phrased it; we are not ready to disclose them in this conversation. She did however point us to a very significant paper given at NANOG 20 by Van Jacobson and two colleagues.

The NANOG paper is titled "Toward Milisecond IGP Convergence." See <http://packetdesign.com/Docs/isis.pdf> > It is by Cengiz Alaettinoglu, Van Jacobson, and Haobo Yu. It suggests that sub-second reroute times would give increased network reliability; support for multi-service traffic (e.g., VoIP), and lower cost/complexity when compared to layer two protection schemes like SONET. Since current IP re-route times are typically in the tens of seconds, the industry should want to do better. There are two choices: replace IP routing with something else like MPLS fast failure recovery or figure out what's wrong with IP routing and fix it.

They prefer to fix the problems of routing. What they are proposing in this paper is a way to decrease the time that it takes changes in routing announcements to propagate within a network from tens of seconds to a few thousandths of a second. If this can be done the operation performance of IP networks will increase enormously. The following paragraph describes some of the paper's conclusions. Because we are summarizing, the arguments that follow are not fully stated. Readers are encouraged to read the entire slide set at the Packet Design URL above.

The Dijkstra SPF algorithm used to compute changes to SPF trees (route forwarding tables) is almost 40 years old. More recent algorithms can compute changes to SPF trees in time proportional to  $\log n$  rather than  $n \log n$ . This allows a net to scale up to virtually any size while bringing the calculation time down from seconds to microseconds. "Consequently stable, robust IP re-routing that works at the network's propagation rate (the theoretical maximum for any re-routing scheme) is both possible and achievable. To get there we have to (in rough priority order): (1.) switch to a modern algorithm for SPF calculation, (2) make the granularity of the hello timer milliseconds rather than seconds, and (3) allow different detection filter constants for link up and down events."

### Exchange Points pp. 7 - 15

We examine the world of exchange points which are undergoing dramatic increases in numbers as propelled by the need of ISPs to peer in as many places as possible to reduce transit costs and by the business opportunities opened by the neutral business model exchange where fiber providers, carriers, ISPs, web hosting companies, ASPs, storage networks, SMTP specialists and others can gather together under one roof to do business with each other in an extremely cost effective manner. Well known players in the neutral exchange point market are PAIX, and AboveNet both owned by Metropolitan Fiber Networks (MFN) and Equinix. MFN's AboveNet has six facilities - three in the US and

three in Europe. PAIX expects to have six exchanges in the US open by year's end. Calling their Exchanges Internet business Exchanges (IBX), only Equinix caters to the full range of clientele listed above.

An article in the November issue of *Telecommunications* points out that "companies such as COLO.COM, CoreLocation, Equinix, Eureka, MFN's PAIX.net, and Switch & Data are expanding into major telecom hubs. Unlike the traditional telecom hotel model where carriers lease physical space from a building owner, neutral COs provide air conditioning, backup DC power, HVAC, dust control and high-level security in addition to real estate." In this respect an interesting new player is the NAP of the Americas. Located initially in a Miami Florida Switch and Data Facility building and managed by Telcordia, the NAP aggregates traffic for a group of carriers which service Latin America.

The *Telecommunications* article concludes that "while demand will not end soon, there is concern that the colocation industry is heading toward a pseudo space glut with plenty of space but nowhere to connect to the backbones. It is estimated there are 42 national CO providers, with more than 25 million square feet coming on-line next year, a 50 percent increase over current availability."

Of the half dozen or so major players Equinix remains the most interesting. It seems to have concluded that it can achieve economies of scale and enough profits to conclude a very ambitious global build out of 30 very large facilities by building new state of the art centers that are on the order of ten to even twenty times larger than those run by its competitors. Equinix is very wisely taking great care not to run into the trap of insufficient backbone connections but building only in locations where fiber from at least five different providers is available. It has raised over \$600 million in capital. It has negotiated a build out schedule with Bechtel that will depend on its ability to finish new exchanges on schedule, populate them quickly with customers and gain adequate cash flow to finance the next buildings on its schedule.

Given the scope of its ambitions it is fortunate for Equinix that it has the greatest technical depth of all the players having been started by the founders of PAIX and having acquired the services of respected infrastructure architects Bill Norton and Sean Donelan. We interview Equinix CTO Jay Adelson and his colleague Lane Patterson. The result is an in depth look at issues such as peering, by pass of ILECs in urban areas, the business models of metro area fiber providers and fabrics for interconnection. Reviewing the myriad of details involved in Equinix's understanding of its industry, the attention that has been given to developing an understanding of the kinds of assistance useful to its customers is quite impressive. An understanding of the economics of interconnection for players of widely differing size assures that a range of options from direct interconnects to access to a variety of switching fabrics is available.

continued on next page

Given the pace of change there are many options open to Equinix customers. The Equinix staff is there to help ensure that the customers derives the greatest synergies possible for his business by making the wisest choices. As Adelson says: "If you are shipping a lot of bits you basically have an economic and technological justification for participating at an exchange."

### **ENUM Administration, pp. 16- 17**

The IETF and ITU concludes successfully a set of meetings in Berlin at the end of October. The meetings resulted in a set of understandings that will keep the ENUM domain out of the hands of ICANN and that pave the way for deployment of ENUM provisioning under the e164.arpa domain with each national e164 numbering administration authorized to choose the distribution entity within that countries boundaries. The likely division of ENUM administration in the US into a single neutral Tier one records keeping entity and multiple tier two provisioning entities is explained. Finally abstracts and URLs of the internet drafts relevant to administration and the Berlin meetings are listed. Meanwhile Network Solutions driven by its business interests ignored its earlier chastisement for the premature start of ENUM trials and announced that it was trailing non ascii character domain names before the IETF completed its standardization work. This prompted a public rebuke from ISOC warning that NSI's action will harm the stability of the Internet Domain Name System. See: <http://www.infoworld.com/articles/hn/xml/00/11/08/001108hnmultilingual.xml>

[www.infoworld.com/articles/hn/xml/00/11/08/001108hnmultilingual.xml](http://www.infoworld.com/articles/hn/xml/00/11/08/001108hnmultilingual.xml)

### **IS-IS Bug Noted, pp. 18 -19**

NANOG discussion of how UUNET's architecture proved susceptible.

### **ICANN as Launderer for DoC Policy. pp. 19 - 21**

On November 13 ICANN posted to its web site a response to the law suit against it stating that it had no authority to add new names to the root after all. Such authority, it admitted belonged to the Department of Commerce. See <http://www.icann.org/tlds/correspondence/esi-v-icann-13nov00.htm>

Thus continued the shell game between the American government and its illegitimate, illegal offspring known as ICANN. We republish a conversation on from the Domain policy list where Jim Dixon shows how Europeans are becomingless willing to accept ICANN authority. Dixon also explains how .eu was created as a country code domain to keep it free of ICANN's control.

### **DNRC Letter to Esther Dyson, pp. 21-22**

The Letter calls on ICANN to drop its clean slate Member at Large Study showing that the "study" breaks a two year long series of public promises by ICANN leaders.

**Battle for Cyberspace-  
How Technology and  
Political issues May  
Affect your Internet  
Venture cost of \$695  
and now available.**

### **Subscription Rates**

Choice of either ascii or Adobe Acrobat (PDF) format 1. Individual; College or University Department; or Library; or Small Corporation - \$250 2. Corporate - (revenues \$10 to 200 million a year) - \$350 3. Large Corporate- Revenues of \$200 million to \$2 billion per year - \$450 4. Very Large Corporate- Revenues of more than \$2 billion per year - \$550 Site License: The right to distribute ascii and PDF via email to all employees of corporation. 5. Small corporate: \$450 6. Corporate: \$650 7. Large Corporate: \$900 8. Very Large Corporate: \$1150 . Site License Distribution via intranet web site \$400 a year additional. See [www.cookreport.com](http://www.cookreport.com) for more detail

Gordon Cook, President  
COOK Network Consultants  
431 Greenway Ave  
Ewing, NJ 08618, USA  
Telephone & fax (609) 882-2572  
Internet: [cook@cookreport.com](mailto:cook@cookreport.com)

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