The COOK Report on Internet

Juniper M40 Out Performs Cisco as New Terabit Routers Start to Come on Stream

Tony Li on Juniper Design & Issues of Intelligent Traffic Management at Core of Stupid Internet

Volume VIII, No. 3, June 1999
ISSN 1071 - 6327

Editor's Note: Tony Li is a Distinguished Engineer at Juniper Networks. He went to Cisco straight out of graduate school and came to Juniper in the middle of 1996. While at Cisco he was involved with router design from 1991 to 1996. We interviewed him on February 26.

COOK Report: I understand that while you were at Cisco, you were concerned about the direction their router development efforts were taking and so you left. If this is the case, what did you do and how did Juniper get started?

Li: Juniper got started because our CTO and founder Pradeep Sindhu observed that the full advantages of ASIC (Application Specific Integrated Circuit) technology hadn't been brought to bear on high end-routers. ASIC technology was readily available and router vendors had used them on low-end router applications, but had not applied it to the high-end. The same ASIC technology, applied to bridging produced the Ethernet switch. It was clear that this was a necessary technology for high-density high-speed hardware.

Pradeep came from Xerox PARC and Sun, bringing with him expertise in ASIC development and in building the high-bandwidth memory subsystems necessary in building a router's buffering subsystem.

COOK Report: Was Sindhu just beginning to do this when you left Cisco or did he start it after you left?

Li: Juniper was founded in February 1996 and I joined shortly thereafter. My introduction to the company was through another founder, Dennis Ferguson, whom I knew through previous contacts at Internet MCI, ANS, and CAnet. Dennis was one of the contributors to BGP and gated in the early 90s. Based on his background and Pradeep's, it was clear that Juniper would have an interesting team, willing to build a high-end router without compromises. The technical challenge was compelling.

Building a High End Router without Compromise

COOK Report: How do you define a high-end router and has the product changed in any interesting way from its original definition?

Li: Today's state of the art in high end router design has many challenging components. For the details on what it takes, look at http://www.juniper.net/leadingedge/whitepapers/backbone-routers.fm.html. This was our initial vision and it hasn't wavered. Along the way our customer base requested many refinements. We were able to include almost all and are very pleased with the results.

COOK Report: Over the last year or two, a number of companies have invested in Juniper and its technology. Who are the investors? Is your technology licensed to them? Can they buy your routers at a discount? How does the relationship between Juniper and its investors work? And finally, how did your development efforts get started? How is your company organized?

Li: Initially Juniper was funded following the classical VC model. Things became interesting when we went to our partners for financing. Our partners at the time were carrier and equipment companies: Ericsson, Lucent, Nortel, 3Com, and Siemens/Newbridge. In addition there were end-user investors, such as AT&T Ventures, the Anschutz Family, and UUnet.

COOK Report: Is it rather unusual to have end-users involved in a startup company in this manner?

Li: Yes, partnering with the customer from both a funding and a technology perspective has been both innovative and extremely valuable. From a technical viewpoint, it has given us the early feedback we needed during both the design and implementation phases of development. From a financial perspective, their investment has helped to cement the partnership, ensuring a mutual commitment to the process of bringing the Juniper M40 to market. This type of mutual cooperation is really the only way to turn out an excellent product in this exclusive, high-end market.

COOK Report: What advantage does an end-user gain from participating in the development process? Wouldn't they tend to be more familiar with the product when it went on the open market because they'd have under their belt all of the experience of working with it during development?

Li: Yes, they gain a great deal of familiarity with the product, but more importantly, they provide influence and design feedback early in the process. Given the long lead time of hardware projects and the focus that a start-

On the Inside:

up must have on delivering the right product the first time, getting this design input up front is invaluable. It avoids costly design flaws that would otherwise hamper product acceptance. The ISPs that worked with us had a significant influence on the design of the router. Obviously that included UUnet, but others also played a significant role. For example, Verio was willing to put in a lot of time right up front helping us spec out what they wanted.

COOK Report: Were there any outstanding lessons learned during the development process, or was it simply the usual slow and steady progress?

Li: The primary lesson we learned was that a well motivated team can bring development techniques and technologies perfected in other portions of the computer industry to bear on the development of networking hardware. There can be an effective merger between networking knowledge and mainstream hardware development. The goal is to leverage tevreything that can be learned from the broader market. We think that such an ongoing symbiotic relationship is necessary to stay on the technology curve.

Running a Very Efficient Routed Core

COOK Report: Let's talk about network engineering. The engineering white paper on your web site indicates that MPLS is the solution to the problems created by an ever-increasing amount of traffic on an old architecture consisting of a network with IP overlaid on an ATM core. Does this situation create the need for a piece of hardware like the M40 Internet backbone router?

Li: Yes, the primary motivator is the sheer growth rate of IP traffic. Without products in this class, ISPs have no reasonable means of continuing to scale. Historically, the backbone capacity of most ISPs doubles at least once every year, and with the increase in the user base and in edge circuit speeds brought about by DSL and cable modems, this growth rate should continue for quite some time to come.

Further, given that the bulk of the bandwidth will be consumed by IP, the efficiency of statistical muxing in a routed core is compelling. The traffic engineering mechanisms available in the M40 further improve the overall efficiency of the solution, giving the ISPs the appropriate tools to best use their fiber plant by placing traffic exactly where there is available capacity. We have another white paper that presents this in more detail: http://www.juniper.net/leadingedge/whitepapers/TE_NPN.html.

COOK Report: Does the MPLS protocol aim to do at the routing level the kinds of network engineering that's done with virtual ATM circuits?

Li: Exactly. MPLS allows us to create a hybrid network that is both connection-oriented and datagram-oriented on the same infrastructure. This allows us to take advantage of the best features of each of these models. The datagram model provides us with excellent scalability and stability and continues to utilize the existing global routing infrastructure. Simultaneously, the connection-oriented model allows us to manipulate traffic aggregates in manageable ways, such as allowing ISPs to perform global optimization by mapping the demand matrix onto the physical topology.

Future developments with MPLS will allow ISPs to use MPLS to provide VPN services, wholesale services, enhanced traffic aggregation, and scalable differential routing.

COOK Report: The white paper states that all routers must perform two fundamental tasks, routing and packet forwarding. It goes on to say that one of the unique features of the M40 router is that it completely separates these tasks because it comprises two independent components, a Routing Engine and a Packet Forwarding Engine. To help our less technical readers understand the importance of this task separation, would you please explain the two functions at a more basic level?

Li: We think of the routing function as the actual exchanges of the routing protocols between routers that go into computing the routing table. The forwarding function takes the routing table and actually switches packets through the system. The other significant function of a router is the management component.

Early commercial routers attempted to rely on the processor based platform by overlaying both the routing and forwarding functions on the same sets of processors. This scaled poorly, because it was difficult to allocate the processor between the various functions in such a way that each function received a sufficient share of time to accomplish its tasks, while allowing functions to absorb the unused time left over by other functions. If the forwarding function consumes too much time, then the routing function starves, which in turn causes the soft real-time protocols to become unstable. Similarly, if the routing function overwhelms the forwarding portion, user packets are needlessly discarded. An analogous situation exists with the management functions. This scheduling challenge is compounded by the need to have a simple and efficient I/O subsystem driven by hardware interrupts.

A better alternative is to simply separate out the forwarding function and then divide the processor between the management and routing functions. Because the latter aren't interrupt driven and don't scale linearly with the bandwidth of the platform, the processor demands are more easily managed using typical operating systems techniques. This division also requires that the protocols remain robust and that the microprocessor technology curve continue to support the routing function.

The forwarding function is challenging because of the scale of the lookup rate and the size of the forwarding table. An IPv4 packet contains a destination address that serves as the key for the lookup in the forwarding table. The forwarding table is extracted from the routing table by precomputing the minimal switching action for each prefix in the Internet routing table. Each destination address corresponds to a single prefix in the forwarding table based upon an algorithm known as 'longest match.' Determining the longest match at full interface speeds with the size of the Internet backbone forwarding table is one of the necessary capabilities of a backbone router.

Scalability of the forwarding table is also key. Currently, there are about 60,000 prefixes in the Internet backbone. This has grown from about 5,000 prefixes in 1991. The ability to support the ongoing growth of the table is a mandatory requirement.

COOK Report: Does this mean that for 60,000 different routes there have to be 60,000 different complete sets of forwarding information because when you change from one location [along the route] to another, there's a unique set of conditions surrounding the packet that must also be changed?

Li: That's correct. Routing computation is unique to the particular location within the topology, with each router computing a full routing table and forwarding table for the full set of prefixes. This computation is also concurrent with the exchanges of routing table information among many neighbors, both local and remote. And the number of such neighbors continues to increase, not only because of the number of physical interfaces found on the router, but also because of the number of virtual interfaces introduced by ATM and Frame Relay. Such interfaces create virtual topologies. This has created new challenges because it forces the routing function to grow with the complexity of the virtual topology as well as with the network physical topology.

Interestingly, MPLS need not share this property. MPLS creates a virtual circuit that is overlaid on the routed topology, but does nothing to deter the continued use of the routed topology. In fact, the routing protocols that control the MPLS virtual circuits remain on the routed topology. This causes the routing function to scale with the physical complexity of the system, allowing a vendor to more accurately design a balanced system.

Routing scales as the product of the number of neighboring routers times the table size. Supporting this scaling process is already a major challenge and needs a tractable long-term solution. Otherwise network equipment manufacturers aren't going to be able to support the routing function, even on the fastest processors.

COOK Report: The information about the network conditions surrounding a packet at a particular point in time is referred to as state information, isn't it?

Li: Precisely. The BGP protocol is normally used to exchange inter domain routing information, which consists of a prefix and related attribute information that describes the path to the prefix, the exit point from the domain, the best way to enter the next domain, and certain policy information about the prefix. All of this is recorded in each backbone router's memory in the routing table, where it is compared against other information about the same prefix. The optimal choices are then selected and become part of the forwarding table.

The data stored in the routing table and forwarding table is considered state information about the prefix, and it will scale with the complexity of the inter domain topology. Scaling this state information is possible at all only because the number of prefixes is growing less rapidly than Moore's law drives memory sizes. And this is the case only because the introduction of CIDR caused the number of prefixes in the backbone routing table to scale as the logarithm of the size of the network, instead of linearly with the number of organizations connected to the Internet.

As one of our customers likes to say, scaling is the only problem. Solve that and everything else is easy.

COOK Report: In other words, we're talking about a process of setting up current road maps within the network—a process that goes on in parallel with the actual packet forwarding?

Li: Concurrency is a necessity. Because backhoes keep the Internet topology in a constant state of flux, the routing function is always in demand. An entry-level requirement in the backbone space is the ability to support this flux without disrupting the forwarding of uninvolved prefixes.

COOK Report: Is this why routing updates are associated with the image of casting a stone into a pond and making ripples that move out from the original point of impact?

Li: That's one of the contributing factors. The additional complexity is that routing is a dynamic, distributed computation that relies on all systems performing adequately to achieve stability. If minimal performance requirements aren't met, such as in a system where forwarding interferes with routing, the system can become unable to sustain the portions of the routing protocol that differentiate between live routers and failed routers. Because this now appears as a system failure, it is interpreted as another topology change. If there are a sufficient number of systems near their upper stress levels, this instability can cascade, possibly resulting in extended service outages that can be addressed only by major amounts of manual intervention.

On a historical note, the separation of routing and forwarding processes isn't a Juniper innovation. It actually first appeared in the ENSS on the NSFNet. While it was an obvious necessity in that architecture, some of the stability benefits were not immediately apparent, and the separation took time to migrate into the commercial router design base. Today it's a widely accepted architectural standard.

COOK Report: Were there any particular engineers who played key roles in the development of the idea of separating routing and forwarding to increase scalability?

Li: There were many contributors, from those who did the initial work on the ENSS, to those who proselytized the architecture to commercial vendors, and to the implementers who eventually adopted the architecture within the commercial sector. All have a great deal of experience in Internet backbone operations.

COOK Report: Are you saying that some engineers became very, very interested in the idea, realized there was some urgency to working on what was an obvious problem, and wanted to work as hard and quickly as they could to make as much progress as they could? Are you inferring that it wasn't easy for engineers to go in this direction if they stayed at Cisco?

Li: No, not at all. For most, the departure from Cisco was for non technical reasons. To be sure, those reasons had some technical repercussions that are still playing out. The architectural separation of routing and forwarding was even making major inroads at Cisco, and they've subsequently shipped this beneficial change.

For us at Juniper, the goal has been to build a very large, scalable Internet backbone router and do it with aggressive technology on all fronts, without legacy software to support. Having the efficiency and focus of a start-up has been beneficial.

COOK Report: Desh Deshpande at Sycamore told me something similar. He said that because he doesn't have legacy equipment to support, he can build devices that will have somewhat more limited functionality, but can be used by people building new infrastructure. And he can do it a lot more quickly than a more established company could. Doesn't this hold true for any technology company? It would seem that the ability to rapidly innovate tends to have a definite drag factor if you have worry about maintaining compatibility with an installed base.

Li: The legacy compatibility issue is a serious one. Expectations are reset with the transition to the start-up, which allows us laser-like focus on our core competencies. For Juniper, it means that we can implement a router and not include support for the many legacy protocols currently required in the enterprise market. I'm not saying that this level of focus is impossible in a larger firm. It's just more difficult to achieve with a broad employee base and a customer base with a more diverse set of expectations.

ASICs as Route Look-up Engines

COOK Report: Was Juniper's goal to take ASIC (Application Specific Integrated Circuits) technology and use it to scale upward a router's ability to forward packets and do routing computations.

Li: So far, ASIC technology is best leveraged in the forwarding function. ASICs are excellent tools for implementing high-speed packet buffering, manipulation, switching, and modification. The cornerstone of these functions is the longest-match lookup algorithm. Through aggressive application of ASIC expertise, we've been able to achieve a lookup rate of 40 million packets per second from a single ASIC. Because of the complexity of the algorithm, it's challenging to achieve such speeds without dedicated, optimized hardware.

One of the painful realities about being in the backbone of the Internet is that it's the location that has the worst-case need for route lookup performance. This is because the longest-match algorithm scales as the logarithm of the size of the forwarding table. Compound this with the fact that the forwarding table is the largest in the backbone and the fact that the backbone is where the largest bandwidths are needed, and it's obvious that this is a worst-case situation for forwarding performance.

We were very aggressive in our use of ASIC technology in this area, and we applied our understanding of the forwarding problem to produce an extremely fast lookup engine. Specifically, there are a variety of different data structures and techniques that can be used to realize the longest-match algorithm. Many of these have been described in the literature. Our task was to combine our understanding of ASIC implementation with our understanding of these alternatives and then optimize the hardware architecture to make efficient use of the ASIC's capabilities. Another complicating factor is that the forwarding table is always in flux, so that it's necessary to be able to update it in place, on the fly and without undue overhead. It was a challenging design task. We have met our performance goals. Doing so would have been simply impossible without our very experienced ASIC team.

COOK Report: Where does the concept of switching fabric fit into the picture?

Li: Internal to any router, there is bandwidth used to transit packets from the input interface to the output interface. The interfaces vary depending on the specific media layer in use. In the general case, this interconnect bandwidth is considered a fabric and can be implemented in a variety of ways. For slower systems, bus architectures are appropriate, while for larger systems, more complex switching fabrics are a requirement.

For a router fabric, there are several key properties. The first is the requirement for any-to-any bandwidth, that is to say that the switch is nonblocking for any noncongestive traffic loading. Asymmetric fabrics require the end user to understand their traffic patterns at very fine granularity, such that circuits can be provisioned to specific portions of the switch. The result is a provisioning nightmare for the end user, even assuming that there is a mapping from the traffic matrix to the switch configuration. A further challenge is that traffic matrices drift over time, requiring reprovisioning to support expected traffic patterns.

For obvious reasons, the switching fabric must be fair. That is, over the long term, any input interface must have an equal share of the bandwidth to each output card. At the same time, the switch must be able to support the full rate of any interface card reaching any output interface card.

The switching fabric must be free from head-of-line blocking. This type of blocking occurs if an input card has traffic for a specific output card but is unable to transmit anything into the fabric because some other packet must be transmitted to a second output card first and switch output to the second card is already allocated. The result is that the input card is unable to transmit into the fabric at its input rate, thereby causing unnecessary congestion and packet loss on the input card.

Another key feature is the notion of fabric speedup. Loosely, speedup is the ability of the fabric to deliver traffic to an output faster than it can be transmitted by the output interface. Speedup is necessary to recover inefficiencies in the switching fabric, to ensure that input interfaces have a very high probability of getting an equal share of the fabric bandwidth over the long term, and to avoid destination collisions.

As you can see, switching fabric design is a complex art, with many intricate design considerations. Fortunately, it is also one of the most studied portions of switch architecture. Juniper's innovation in its switching fabric is in an increased efficiency and ability to use less sophisticated technology to implement the same size fabric as compared to other, traditional crossbar designs. The increased efficiency results in a lower part count and parts cost, which in turn results in an improved interface and bandwidth density for the system. The innovation here is in the exploitation of a sweet spot within the design space.

COOK Report: Elaborate on what you mean by a "sweet spot".

Li: There are trade-offs that you have to make in doing any design. One variable was how much bandwidth you could put across the backplane of a system using a particular connector and ASIC technology. We examined the design alternatives and found an interesting and efficient solution. That efficiency translates into fewer parts and a more efficient utilization of the parts you have. This ends up costing less to the end-users because it's less costly to produce.

This is possible because at certain levels of ASIC technology, the number of transistors used actually doesn't change the price of the product appreciably. The marginal cost of an additional transistor actually declines as you scale up the ASIC. Thus, adding additional complexity to the ASIC design to produce a more efficient system can result in a net win. You certainly pay for this in additional development costs, but the return easily justifies the additional expense.

COOK Report: What are some of the difficulties involved in designing ASICs? Is it acquiring the engineering talent? Are there tools like templates and existing software that can assist in the process? Is hardware and software support readily available?

Li: High-end ASIC development is part art and part science, and the networking industry hasn't yet taken full advantage of the capabilities of the technology. Certainly acquiring the talent is a challenge, but after that, all the necessary development tools are available in the marketplace for a premium price. The computer industry has driven the creation of ASIC development methodologies, so the tools are available and the talent exists. We were able to draw on the cream from this talent pool and then work with them on the more specialized problems in a router's subsystem.

By leveraging this talent, we were able to apply state-of-the-art ASIC technology to the system while still producing a system with a reasonable risk profile. And by giving our designers the liberty of architecting the entire system, they weren't constrained by existing off-the-shelf networking parts or conventional networking industry design practices. In effect, we were able to step back and view router design from a fresh perspective.

COOK Report: Have the design processes been refined to the point where you've been able to recoup your capital investment in the design tools, the equipment, and the engineering talent?

Li: Much of our funding went into financing our ASIC development process, which is certainly capital intensive. Fortunately, this expense is amortized across the lifetime of the product, so it's a small investment considering the additional value that our ASIC design brings to the system.

In addition, to contain the ASIC design and manufacturing costs, we partnered with IBM as our ASIC foundry. By working with them and their excellent processes, we were ensured that our design was executed in a timely and efficient manner, with conservative design rules so that we had very high confidence in our designs working without significant revisions. That confidence paid off, giving us a fully functional chip set very early on. That allowed us to bring the system in on time and on budget.

COOK Report: So IBM takes the inputs from you and puts them through their manufacturing process. Is this what you have to do to get a prototype chip, too?

Li: We provide them with a completed design that matches the design rules of their process and they perform their processing, returning completed chips to us. ASIC prototypes are very expensive because they're basically doing a one-off run of an entire wafer. Of course, once the chip enters production, costs fall rapidly.

Basic Functionality of M40 Architecture

COOK Report: Let's look more closely at the M40 architecture by having you describe the path a packet takes through the M40 router. [A series of very good technical papers may be found at http://www.juniper.net/leadingedge/default.htm].

Li: The entire story starts when a packet enters the system through one of the physical interface cards (PIC) that is specific to a particular media (ATM, SONET, etc.). The physical interface card removes the packet from the media framing. The packet is then handed off to a flexible PIC concentrator, which determines the packet's protocol type (e.g., IPv4) and extracts key information from the protocol headers so the router can do a lookup. Our focus is on IPv4 so the primary part of the key is the destination address.

The flexible PIC concentrator then injects the packet into the distributed shared memory that acts as a system wide packet buffer. Some buffering is a necessity because a router must cope with small bursts of congestion without packet loss. Buffering allows us to absorb congestion up to a full round-trip delay time, and this has been shown to give the best performance for TCP connections.

Providing adequate bandwidth into this buffer subsystem is one of the key innovations in the M40. What's interesting about the M40 in this regard is that we've actually created one large buffer that's used system-wide and that's responsible for all packet buffering within the system. This is significant because other architectures that have multiple bottlenecks within the system have to have buffering subsystems throughout the architecture. For example, other architectures have input and output buffers. The result is more memory and more expense.
\

COOK Report: In other words, you need to buffer something only when you can't process it at the appropriate moment. So the M40 router has a generic system wide bin that packets are thrown into whenever the need arises. And an instant later, the packet can easily be identified and taken out of the bin.

Li: That's correct. We have a system wide buffer with fast access times. Very simply, the packet arrives on the system and is placed in this buffer while we perform the route lookup. When the lookup completes, we place the packet in a queue on the appropriate output interface. When the packet eventually reaches the head of the queue and we need to transmit it on the output interface, we retrieve it from the central buffer.

COOK Report: Is all this done within a special ASIC or with some other hardware?

Li: Because of the complexity of the entire process, it's done in multiple ASICs throughout the system with differing parts in different ASICs. This is the reason the M40 had to be designed as an integrated system. It couldn't have been designed as individual piece parts. Simply buying off-the-shelf components from component manufacturers would never have given us the flexibility required to place the functionality where it was needed to execute the architecture.

Again, this is possible only through the application of leading-edge ASIC technology and a systems approach to the overall design process.

COOK Report: Alan Huang (Feb. 1999 ***COOK Report***) has a very different idea about building a high-end router. He claims it's possible to do so using generic off-the-shelf parts and chips. What's your reply to that?

Li: I believe you can construct very interesting routers out of piece parts and even other routers. However, to do so becomes very inefficient and that translates into untoward expense. Huang is looking at basic well-known switching architectures and there really aren't any surprises there. What he has done is to extrapolate from these basic architecture designs to ones using very, very large piece parts, or in other words, existing routers. However, to build a single interesting router you have to be much more efficient than that. Certainly, to hit a price point where customers are interested, you're going to have to be much more efficient.

For Juniper to achieve that efficiency, we've had to go all the way to designing our own low-level ASICs and getting the subsystem fine-tuned to the point where it's actually doing exactly what we want in a very careful manner. The interesting part of Huang's work is that all the architectural principles he points out in his paper, multistage switches for example, are the kinds of things that appear in low-level architectures.

COOK Report: Let's return to the topic of following packets through the M40.

Li: I left the packet narrative at the point where the packet is injected into the buffer by the flexible PIC concentrator. Once the packet is in the buffer, the flexible PIC concentrator is responsible for extracting the IPv4 destination address and any other fields we're going to do a lookup on. This information is bundled together in what we call a key. A key can be quite complex, especially if we're doing other things like multicast. The key is transmitted to the Internet processor ASIC, which actually performs the forwarding lookup. In addition to doing the lookup, the Internet processor also extracts information from its forwarding table regarding where the packet should go. This information is passed to the outbound flexible PIC concentrator. It's then placed in a queue for the particular interface on which the packet exits the box. Once the media becomes available on the output side of the box and the packet gets to the head of the queue, the packet is extracted from the systemwide buffer and transmitted out at line rate.

COOK Report: How many Internet processor ASICs are there per box?

Li: Only one. It handles all interfaces in the system including any type of input/output media like ATM, SONET, Frame Relay, etc. And it handles them all simultaneously at line rate. By centralizing this function, we were able to remain highly efficient. It means that we pay once for all our forwarding table memory and forwarding hardware. If you distribute forwarding lookups across more and smaller parts in order to scale the forwarding function, it becomes much more expensive.

COOK Report: Is it the system wide buffer that allows the great variety of input/output media to be maintained as a smooth flow in and out of the single Internet ASIC processor?

Li: Backbone traffic is always bursty, and the Internet processor was simply designed with this traffic distribution in mind and without compromising system performance. Again, centralizing the buffer helps us by minimizing the amount of memory in the system while retaining the bandwidth-delay product in memory capacity necessary for buffering across all interfaces. This makes the buffering subsystem very efficient, thereby minimizing the amount of memory and in turn minimizing the cost.

COOK Report: Can you compare figures regarding memory size in the M40 and other routers of equivalent power?

Li: In general, our router memory is a factor or two smaller than that of conventional routers because these routers have to do both input and output buffering. The double buffering is necessary in other systems because the switch fabric can become congested, forcing traffic to queue up prior to traversing the fabric. By unifying the buffers, we were able to avoid this division and thus avoid unnecessary replication.

COOK Report: What role do the line cards play in the M40?

Li: A line card is a combination of a flexible

PIC concentrator and up to four PICs. A fully populated M40 system provides 32 PIC slots—an industry-leading port density of one slot per rack-inch—which offers complete mix-and-match flexibility for line card installation. Because the switch fabric has a significant amount of speedup, all line cards operate at wire rate for all packet sizes.

COOK Report: What are the components of the M40 Packet Forwarding Engine?

Li: The Packet Forwarding Engine refers to the Internet processor, all the attached flexible PIC concentrators, and the physical interface cards. In short, all the hardware in the forwarding path is considered to be part of the PFE. And we distinguish the PFE from the Routing Engine, which is actually the computer we use to compute the forwarding table.

COOK Report: Say something more about the routing engine and how it communicates with the PFE.

Li: To begin with, there's absolutely no hardware innovation necessary for building the Routing Engine. We built it using all off-the-shelf components. The interconnect between the PFE and the Routing Engine is based on straightforward interprocessor communications protocol. The Routing Engine then has the responsibility for running the routing protocols, computing the forwarding table and then placing a consistent forwarding table within the PFE. As I mentioned earlier, the forwarding table is under constant revision, so this is a constant, ongoing process.

COOK Report: Your ASIC designs seem to be quite an accomplishment. But technology changes quickly. How much room for growth does the Internet processor ASIC have? Will it last for a couple of years or longer? When will there be a processor that can handle 80 million or 400 million lookups per second? Or is this kind of growth in lookup capability handled by MPLS?

Li: We know that ASIC technology is going to continue to progress, roughly according to Moore's law. We hope to leverage that continuing evolution along the technology curve. Certainly, the existing 40-million-packet-per-second performance of the current Internet processor will be topped at some point. But by continuing to apply the latest technology, we believe that the Internet processor can continue to track the ASIC performance curve.

At the very least, we can look to IBM's new SA-27 technology for an obvious improvement. SA-27 technology was brought out last year by IBM and incorporates copper into the ASIC, replacing the use of aluminum interconnect. This has the effect of making the ASIC faster and smaller. Placing our design into this technology is relatively easy thanks to a consistent tool set across different ASIC technologies, so little additional design time would be needed to create a faster chip and scale up forwarding lookup performance.

MPLS really isn't necessary for fast forwarding lookups. It's true that MPLS does make the forwarding lookup easier in some sense, because the lookup algorithm is now a table lookup and the key itself is shorter. However, it's still necessary and possible to do normal IPv4 forwarding lookups at line rate. So MPLS doesn't provide significant benefit in forwarding performance. Its advantages lie elsewhere, namely in its ability to engineer traffic to a particular topology.

COOK Report: Doesn't that have applications all across the board as far as memory chips are concerned?

Li: Yes, improvements in the base technology will continue to help improve performance and scalability. The key issue here looking forward, however, is that the demand for Internet backbone bandwidth continues to outstrip Moore's law and the rate of improvement in base technology. This implies that future developments will require more effort and more thought than simply applying the latest technology. That will certainly be necessary, but not sufficient. One can imagine higher degrees of replication, new architectures, and the application of new technologies. For example, at some point, it might become necessary to design fully custom chips, a technology that is reserved for microprocessor design today.

Routing Software and QoS Issues

COOK Report: Now that we've covered some ground in hardware, let's get back to software. Tell me about the MPLS protocol and the role it plays in the M40.

Li: MPLS is the current state-of-the-art in the Internet routing and forwarding architecture today. The primary significance of MPLS is that it allows you to change the basic forwarding architecture of the Internet. That, in turn, allows you to change the basic routing architecture, which allows you to perform an entirely new set of functions that will in turn enable an entirely new set of services. As I mentioned before, MPLS gives us a hybrid network architecture, where we can support both datagram mode forwarding and a connection-oriented service in parallel. It allows you to use this hybrid strength to route packets through the network based on something other than the destination address. This gives the network an incredible degree of increased flexibility, which will allow us to provide new services such as voice over IP and VPN services on an integrated Internet backbone.

MPLS can also be used as an aggregation mechanism for traffic that would otherwise require a very detailed analysis of packet characteristics before performing a specialized forwarding function. By aggregating this information, we can limit the amount of the state in the backbone and maintain reasonable scalability properties there while still providing special services at lower aggregation points.

COOK Report: So with MPLS you could have many flows at the edge of the network and each of the flows could be given various definitions of QoS? But when the number of flows increases into the hundreds of thousands along the core backbone, isn't a new problem is created? —Namely that there's a conflict between the amount of separately defined QoS levels you'd like to have for customers at the edge of the network and what you can actually aggregate and deal with by routing on the backbone.

Li: Correct, the amount of state information (memory used per flow) in the backbone would be unmanageable without some form of flow aggregation. At certain interconnects, we've seen about 40,000 new flows per second, for example. Trying to deal with this using any of today's technology on a per flow basis is simply unworkable as the number of flows per second is likely to continue to scale with the growth of the Internet. Because this is well ahead of Moore's law, this might be possible in the very short term, but will certainly lead us down a dead-end path in terms of scalability.

There are currently two basic possible approaches that would seem to have a better scalability story at this point. The first is the application of MPLS to aggregate flows with QoS and routing properties. The second approach is the work that is being done in the Diff-Serv working group in the IETF. Diff-Serv provides a convenient and easy way of "coloring" packets with particular levels of QoS requirements. These are bulk, generic kinds of "coloring" that are not flow specific and thus require minimal amounts of state within the forwarding function in a router.

Both of these approaches would suffice and it's reasonable to expect that both of these will be used in different parts of the network. Diff-Serv is appropriate for simply communicating QoS information and is sufficiently light weight that it's reasonable to expect hosts to participate. MPLS can effectively encode the same information and couple it to other properties within the forwarding architecture, providing both a QoS function and traffic engineering functions, for example.

COOK Report: How many different colors have been agreed upon?

Li: The current Diff-Serv proposal supports up to 64 different "colors", or more precisely, 'Per-Hop Behaviors' (PHBs). Some of these are defined to be global, some are reserved for local use. It's left to each provider to determine the PHBs that are applicable for the services that they wish to deliver. However, because the definition of a PHB can in fact be wholly local to a provider, it raises an interesting question about the ability to define and deploy interprovider Diff-Serv functionality.

The downside to Diff-Serv deployment, even locally, will be a multiplier in the management and state overhead within the network. For example, to support a PHB, it will be necessary to enable the appropriate support parameters and characteristics throughout a provider's domain. This functionality is likely to result in traffic policing, accounting, billing and increased traffic engineering effort throughout the network. The constant multiplier is the number of PHBs supported by the domain, so one might reasonably expect that a provider will be very careful about the number of services it will actually support.

With interprovider Diff-Serv deployment, the situation becomes even more interesting. Without a clear global service definition, the provider is forced into a series of bilateral agreements, trying to match a service level with other service levels as defined by other providers. When multiple providers are in the path, the legal situation might quickly become intractable.

Even if there is a simple technical mapping between service levels, there might also be interesting financial issues. Two providers with equivalent service levels might not be exchanging traffic in a symmetric way. Because QoS traffic intrinsically requires a premium, there will need to be some agreement to divide this premium between the originating provider and the destination provider. The situation becomes even more complicated if there are one or more transit providers. There would need to be some division of the premium between originator, destination, and the various transits. This division would need to take into account the value provided by each of these providers. These values are not necessarily equivalent. One provider, for example, might provide a metro transit service. Another provider might provide transoceanic transit. A rational system would result in differing shares of the premium for different providers, according to the value added by each.

Settlements for best-effort traffic have been proposed before, but never deployed. QoS would further complicate the situation in that it would place the burden of security on the billing system and on network access. Such security is necessary because differentiating traffic opens the danger of theft of services through misclassification of the traffic.

COOK Report: One of the most intractable problems with the current Internet business model is the desire to have various applications accompanied by service level guarantees, which implies various types of QoS. But for the reasons you just mentioned, when service providers' boundaries are crossed, one provider might have three colors of QoS and the other might have ten colors. It's unlikely that all providers will offer the same colors of QoS because the capital investment might vary from one provider to the next and providers are going to want to use QoS to differentiate their services. What impact will this scenario have on your hope that traffic will be able to cross provider boundaries with something other than best-level service? It sounds to me like you might be chasing the end of the rainbow that always appears in front of you but is never really reached.

Li: This is a very challenging policy question and unless there are clear common denominators established, there is little hope of deployed inter domain QoS. On the positive side, it's very likely that the demand will be for Internet-wide QoS services. Differentiated services within a single provider's network is of interest to the provider, but to the end user, who wants to use the Internet backbone as a global facility, differentiation with a limited scope isn't a practical solution.

An Internet Business Model, for Inter Domain QoS

Such a limitation might be possible if the customer is interested only in a limited service, such as a small extranet in which all parties can be coerced onto a single provider, but for true Internet applications, this is clearly intractable. The extranet application can work because a single provider can provide the centralized control over the service, but there's no obvious mechanism to extend this to a global service.

COOK Report: You've hit on one of the most critical issues involving the Internet business model. What are some of the more interesting solutions being put forth? Who is most prominent in this policy area? Where is the work being done? Do you have a sense yet of where all this is headed?

Li: It's certainly a critical business issue and needs to be adequately addressed. The IETF is not a good forum for dealing with business issues, so there's something of a vacuum for resolving this. There is some work happening within the IETF to address some of the technical definitions of certain forwarding classes, but this is not the same thing as having a clear and common service definition. It's surprising that more progress hasn't been made here, but my impression is that providers will start by using the technology in their own networks before attempting to interoperate. A difficulty with this approach is that it might well leave various domains with disjoint sets of services, making it that much harder to deploy a uniform set of services after they already have an installed service base.

Another concern is that some providers are expecting to differentiate themselves based on their domain-specific services and are not interested in supporting globally defined services. This would present a drag on the deployment of global services that could be overcome only when there is a de facto standard and the non participants are the exception.

It is possible that these scenarios can be avoided, but it will take initial experimentation with interprovider QoS agreements, in both bilateral and multilateral agreements. An appropriate forum where such issues could be aired and global service definitions discussed without antitrust problems is needed.

COOK Report: Does it make sense to say that the public Internet might have to remain a best-effort service network? But if you look at virtual private intranets, isn't this where you'll get QoS? But is QoS still going to drive the largest players into becoming worldwide in nature because even with virtual private intranets, at some point you want to communicate with someone who might not be on that intranet, right?

Li: That's another problematic scenario. It's particularly troubling but is a distinct possibility. It implies that anyone wanting global QoS for their global intranet is forced to use a worldwide provider. That scenario is even more challenging for extranets, because all participants would be forced to connect to a single provider. It becomes a phenomenal vendor lock for the provider. Further, given the barriers to entry in achieving a global footprint, it's not clear that you would end up with meaningful global competition anyway. The result is a virtual worldwide monopoly.

The bright side here is that this might drive some of the smaller players to collaborate to prevent exactly this type of scenario from playing out. Unfortunately, it hasn't happened so far, so we're stuck with best-effort service. Of course, if the true actual service level associated with best effort traffic suddenly and miraculously improves, it's possible that this could kill the demand for specialized QoS services. Past history with the Internet, however, doesn't make this seem likely.

COOK Report: To what extent are people trying to solve some of these service issue dilemmas by building overlay networks? I'm thinking of AboveNet and Enron. For instance, Enron is using its network only for aggregation and content delivery, which says, in effect, that they'll connect the largest content producers with as many of the ISPs as possible. And instead of dumping their worldwide web content through the public Internet, they dump it onto their overlay network and then offload it in the 25-50 or 100 POPs around the country. This approach might buy you a little time but it doesn't really buy you a solution, does it?

Li: Just covering the continental U.S. is not nearly sufficient. If QoS continues to be important to some users, it's likely that it will be provided by players who are already international in scope. Look at the simple application of building an Internet-based VPN for an organization with international field offices. In today's global economy, it doesn't take a very big company before you start to establish overseas offices, so this problem isn't confined just to the Fortune 500 companies. The overlay approach, or more accurately, a bypass approach, might aid in solving a limited set of problems, but it is clearly far less flexible than a ubiquitous set of QoS services.

COOK Report: Some people are trying to throw bandwidth at the problem and use different lambdas to provide QoS, but there's going to be a limit to that. Eventually you run out of bandwidth to throw at the problem. One wonders if McCaw is thinking about a global bandwidth solution. Next Link has purchased about $700 million worth of Level 3's fiber. Craig McCaw has an interest in companies like Teledesic. It might be intriguing for him to have a global network where, if you had a temporary bandwidth shortage, you could switch bandwidths from terrestrial to satellite or vice versa. Would you like to comment on all this?

Li: Throwing bandwidth at the problem is a fine strategy to delay providing QoS, but it will never be a long-term substitute unless you can continually feed the bandwidth monster. While it might prove to be possible in the short term thanks to one-time bandwidth multipliers like DWDM, it's by no means certain that anyone will ever be able to provide sufficient bandwidth. And perhaps more importantly, it's not clear that in the long term, selling only best-effort service while paying exorbitantly for scaling bandwidth will prove to be the market winner.

I do expect that all the interesting players will continue to support best-effort traffic in the long run. It would seem to be necessary to have that as a service offering so that economy services can be delivered alongside premium services and result in a rational bill. And we can see that to continue to support best effort-traffic, the provider will certainly have to continue to scale bandwidth rapidly. For this reason, I expect that facilities-based providers will have considerable leverage in holding the best-effort services and thus the vast majority of interesting traffic.

As to utilizing satellite bandwidth, there are some interesting technical challenges involved there. As you know, satellite bandwidth comes with a penalty of additional delay. While TCP can be run over satellites with good throughput, most common implementations aren't prepared to support this increased delay and so will deliver decreased throughput. Interactive applications, including Web surfing will be adversely affected regardless of the implementation.

OS Software

COOK Report: The M40 router uses the JUNOS operating system. People are saying that everyone knows how to deal with Cisco's operating system, but now they're going to have to learn a new one if they use Juniper routers. What's your response to that?

Li: The ISPs we've been working with have been pleasantly surprised when it comes to JUNOS. What we've designed is certainly different than IOS, but it is not so radically different that it's difficult to learn. Certainly there are some syntactical differences, but the semantic commonality plus some of the usability features that we've built into the command-line interpreter make the syntactic differences easy to work around. Furthermore, because we're focusing on the high-end ISPs, our user base consists primarily of highly experienced senior engineers who are already adept at manipulating multiple operating systems. They've found that they can easily acquire JUNOS proficiency in a few hours. Some are even at the point where they prefer it to IOS, so the small learning curve isn't presenting a significant issue.

COOK Report: It's my understanding that the M40 routers are the only ones to perform at OC-48. Would you say something more specific about that? Even though you're in front of the pack with this OC-48 feature, just how large is the market for huge backbone routers on major backbones? Will the size of the market limit your growth opportunities? What effect will this have on the manufacturing process and issues like economies of scale?

Li: Currently, the M40 is the only installed system that supports line rate OC-48. There are several others who are close behind, so this situation won't last too much longer. Fortunately, hardware performance is not the only benefit of the M40. The bandwidth density, interface diversity, and software maturity are also unmatched by any competitor. As to the size of the market, we've seen numbers suggesting that the Internet backbone market, including the telco equipment, can become a $20 billion market. That's sufficiently large to keep Juniper busy for quite a while with any reasonable market share. If and when we do address other markets, the increased economies of scale can only benefit our manufacturing processes. Because we've partnered with Solectron, a large well-known manufacturer of networking equipment, we'll be able to ramp quantities as needed.

Scope of IP Telephony Protocol Development Expanding and Speed Slowing

We asked someone involved in IP telephony protocol developmenty for his views on the state of the field after the Minneapolis IETF meeting in March. He felt more comfortable being off record because of some of the political sensitivites involved.

On those conditions he told us that almost 1/3 of the attendees were involved in IP telephony work groups. However he maintains that it is now clear to him that the protocol process will take much lonmger than seemed likely last summer. Why? Because it is no longer a case of IP innovators linking and meshing with the PSTN. Rather the carriers themselves have decided that the economics of IP over DWDM networks are so compelling that they are adopting IP telephony. Unfortunately their adoption means that IP protocols that duplicate all the features of the old intelligent network must be developed. 800 numbers. 900 numbers. Billing systems. Number portability. the whole nine yar ds. This is a very complex task. It is likely that these developments will strain relations between the IETF and the ITU because now the IETF is for the first time developing the very standards that, apart from sprectrum management, are the very raison d'etre for the ITU and ETSI. All this is happening against a backdrop of the victory of SIP. People seem ready to abandon the idea that this development work can be done on the foundations of alledgedly interoperable H.323 standards. Perhaps the most popular T shirt being worn in Minneapolis was one proclaiming the day that SIP became an IETF standard.

MCI's vBNS Service Explores QoS

Rick Wilder Reports on Testing of Juniper M40s and on Future of Bandwidth Intensive Applications

Editor's Note: Rick Wilder was at Mitre from the mid-1980s to early 90s where he did protocol research and development for TCP/IP and OSI packet networks, particularly in the areas of congestion avoidance and control. From1992-94 he worked for Advanced Network Services (ANS), which, at the time, was operating the NSFNet backbone. There he worked on plans for QoS services which really never saw the light of day in that environment, and a secure encrypting router for building the vBNS on the Internet. He came to MCIWorldCom primarily to put together Internet MCI. As it turned out, he had worked on the vBNS proposal while he was at ANS, and the NSF award went to MCI just after he landed here. As a result, he was the one who started up the MCI engineering crews for the vBNS in 1994 after MCI did the initial rollout of Internet MCI.

COOK Report: Can you comment on the idea of using the vBNS to test a range of routers?

Wilder: Early on, in 1993 the vBNS was different than it is today. It started out as a supercomputer center interconnect. In that particular kind of environment, we had two different kinds of routers because we had a very important requirement for moving traffic from the hippi infrastructure in the supercomputer centers out over an IP backbone. At the time, the only off-the-shelf router that could do that at a high data rate was the Net Star gigarouter. Net Star was bought by Ascend and the Ascend GRF is the descendent of that Net Star gigarouter. We still have those GRS routers in the network at the supercomputer centers.

Why do we use different routers? We've never had a philosophy dictating that we had to use either different routers or all the same routers. We prefer using different routers simply because one of the goals of the project is to introduce new technologies. And you can do more types of exploration if you're using multiple platforms.

COOK Report: So what are you doing in California and what are the kind of problems you're trying to solve there?

Wilder: We're basically in the prototyping stage of the next-generation backbone. We've gone through this already a few times with the vBNS. We started out with IP over ATM architecture for OC3 trunks and ran through a couple of different architectures for OC12. Now we have a prototype for the technologies we're going to use for OC48.

Two Flavors of RSVP at OC12

Let me say a little more about OC12. The current architecture for the OC12 trunks is IP over ATM over SONET. We finished the deployment of this architecture nearly 3 years ago. We're actually in the process of deploying RSVP at this point. It's not a production service yet but will be as soon as the router software is stable. The result will be a reserved bandwidth service where we get RSVP signaling for reserved bandwidth sessions from the IP end stations. When those RSVP requests hit our routers, we map them into ATM switched virtual circuit signaling. When we set up a reserve flow, we signal an end-to-end path through the ATM infrastructure and have the routers map the traffic for that reserved flow onto the correct virtual circuit.

COOK Report: Is this kind of mapping onto ATM switched virtual circuits something that's been created within the past 12-18 months or has it always been around?

Wilder: We actually started out with this idea with Cisco exactly two years ago and we've continued to work together with them. They've put some new functionality into their IOS to support us being able to do this service. Meanwhile, the world of standards has thought along these same lines and there's currently an Internet draft on how to map RSVP traffic descriptions to ATM layer signaling.

COOK Report: Do you know if RSVP is being currently used anywhere in a production capacity or internally in some large networks?

Wilder: I'm not aware of its being used today in a production environment on any large scale. Keep in mind that RSVP comes in two different usage flavors. The first is the end-to-end bandwidth reservation model I just described. This model isn't of widespread interest to public Internet providers because it doesn't have the right scaling characteristics and doesn't fit the need to go across different networks. But if you're in a virtual private Interent, it's a different picture. The situation we're in with the vBNS, as opposed to the public Internet, is that we have a small number of large bandwidth flows. So for that small community of users it's possible to set up enough high bandwidth flow reservations to satisfy the class of special applications that need them without running into big scaling problems on the backbones.

The second flavor of usage is relevant to commercial Internet backbone. They are using RSVP setting up label-switched paths through the backbone. The application here is traffic engineering. (I'll have more to say about this type of usage and RSVP later in the interview.) Right now, as you know, a lot of the big Internet backbones use an ATM core interconnecting their routers. So it's a layer 2 interconnect between the layer 3 devices. Therefore, they're able to use the placement of virtual circuits at the interconnect as a way to tie the right amount of traffic flows to a given physical path, i.e. to load up the path to the right level without overloading and causing congestion.

As traffic conditions in the network change, it's possible to easily change the loading of paths on the fly. However, my experience has brought me to the conclusion that how quickly the changes need to be made and how dynamically the mapping of flows to the backbone needs to be, varies with the scale of the backbone. Suppose you have really large numbers of flows multiplexed together on very big pipes. For instance. OC12 or OC48 pipes carry hundreds of thousands or millions of Internet flows per minute. At this level of aggregation, the physical properties of the traffic are such that they don't change very quickly and you probably don't want to restructure the traffic engineering very often. So in this situation you really don't have to make quick decisions to reroute traffic.

COOK Report: What you're saying is that if one of those flows is arbitrarily New York to Chicago or New York to Los Angeles, it might represent anywhere from 500 to 2, 000 different significant users. But if one of the users ramps up their bandwidth, there's a good statistical probability that another user will ramp down during the same period.

Wilder: Right. The characteristic of the behavior of a big aggregate is that it is much more steady than that of behavior of any given user.

What we're doing right now is based on ATM VCs as a way to map the layer 3 traffic onto layer 2 paths. Looking toward the future, however, I don't think that's going to continue to be the best way to do it. The reason is that cell switching and very high data rates are hard on router interfaces where you have to break the packets up into cells and put them back into packets again on the other end. (Some refer derogatorily and even contemptuously to this process as packet shreading.) I have to admit that this process does use up bandwidth for the ATM headers, but as time goes on this may become less significant as the cost of bandwidth goes down. What I think is more significant is the amount of complexity in the router interfaces needed to do the segmentation and re-assembly, i.e. the conversion of packets into cells and back again. The complexity is actually in the hardware and/or firmware and this, in turn, requires that the interfaces themselves be more complex and more expensive. For example, when we went to OC3 routers based on ATM, it took longer to get reliable router interfaces than it did with DS3 interfaces or with packet over SONET interfaces which is what we're currently running..

Segmentation and re-assembly is typically done with chips on the interface cards. If you look at Internet capability in the last few years in terms of scaling, you'd have to say that one of the significant things is that getting good OC12 interfaces out into the field was delayed by having to have those shared chips, the segmentation and re-assembly hardware, in those cards. In fact, Net Star was very much delayed and put behind the technology curve by choosing to do only ATM at OC12 rates and they ended up having to wait a long time to get good shared chips. If they had done packet over SONET, I think they may have been able to take advantage of a very nice internal architecture they had in the box. But they were hamstrung by the parts availability for the interface card.

Packet over SONET

COOK Report: Could you say something more about packet over SONET in the vBNS?

Wilder: Because of the shared problem I've just been talking about, we're going to be pushed toward using packet over SONET as the transmission technology and using something like MPLS, at least initially, to replace the traffic engineering capability we currently have with ATM. If we do all of this, we're in effect replacing ATM in the core of the network. However, I think ATM will continue to be a very useful access network technology. But for the very high bandwidth backbone trunks (OC48 and before long OC192), it's going to be more attractive to do packet over SONET, packet over WDM or packet over some optical technology.

And I'm not sure that gigabit Ethernet framing is going to be the choice. There are several things in the works out there right now. In fact, SONET framing has a bit format for high speed trunks and that's not necessarily a bad thing. It can handle the OC48 interfaces we're dealing with, for example. The chip sets are available and the overhead isn't very high. I think it's about 3-4% which isn't a killer by any means. So I really don't have a problem with the framing format. The problem with SONET as a technology that some people have is related to the price of interfaces and the management systems, and the total price of putting in the transmission infrastructure with SONET support. It may turn out in the years to come that there's a less expensive way to provide data transport at this very high bit rate.

COOK Report: What's the next step that will take you from ATM to OC48 and Juniper routers?

Wilder: What we're doing with the OC48 next-generation backbone and Juniper routers is essentially replacing the ATM core of the network with MultiProtocol Label Switching (MPLS) on the routers. MPLS will be used to set up label-switched paths in place of ATM Virtual Circuits in order to carry the layer 3 traffic. RSVP is likely to be the signaling that will be used internally in the core of the network to put these label-switched paths in place just as ATM level signaling is used to put virtual circuits in place in an ATM network.

MultiProtocol Label Switching

Before continuing, let me say a few words about label-switching. A label switched path is an MPLS term for the MPLS equivalent of an ATM virtual circuit. It's a very efficient way to forward traffic because it forwards full packets rather than cells. The actual forwarding isn't much different than the way you forward an ATM cell in that it has a label or tag associated with the packet used to select the label-switched path it's going to take. It works exactly the same way a virtual circuit identifier works in ATM. MPLS is the term used by the IETF working groups that are doing standardization of MPLS. Tag switching came slightly earlier with Cisco's first implementation of this kind of technology. (Tag-switching was Cisco's name for its implementation of MPLS. MPLS is the generic term.) Cisco made their implementation description available to the IETF as input to the standardization of MPLS.

There are now multiple working groups in this area looking at specific aspects of traffic engineering and how the forwarding will work. There's quite a high degree of interest in this area and it's been going on for more than a year now. I don't remember the exact date when all of this got started.

COOK Report: Since we're clearly in the pre-standardization phase of MPLS, what exactly do you mean when you refer to "an MPLS implementation?" Do different companies have different versions of the implementation? And does anyone have a real production level implementation yet?

Wilder: We've very much settled on the header format which is a 32 bit header that contains the tags and a little bit more information. What we probably don't have interoperability with yet is the signaling involved to set up the label-switched paths. This is going to require more work.

As far as a real production level implementation goes, it depends on your requirements for MPLS. You can reliably set up label-switched paths between a set of Cisco and Juniper routers, for example, and it will work. But I don't think we have all of the traffic engineering support we'd like to see by any means. We're working on this with both of those vendors.

At this point we wouldn't run a huge production backbone with MPLS on the backbone of UUNET or Internet MCI. Within the next several months, however, we're likely to do just that in the vBNS. We can do this because the vBNS is a relatively small scale network. UUNET would have a bigger set of requirements and a little more stringent testing required for use in their very much larger backbone. Keep in mind that UUNET handles its huge amount of traffic with separate OC12 trunks. I'm sure they're looking at doing OC48 very soon. While there are no significant limits today to how many OC12s you can tie together, there are problems you need to deal with in terms of the layer 3 routing protocols and the virtual circuit counts at each switch and interface. If you look at an ATM switch, it has a limitation on the number of VCs per interface in the switch and on the total VCs for each switch. You have to be careful about these things.

COOK Report: If you use OC12s and packet over SONET and not ATM in the core, does it become a different matter?

Wilder: In that case you have the scaling properties of MPLS and label switched paths. You can choose to use that instead of ATM. We don't know enough about this scenario but it has the potential to have better scaling than ATM. But it's part of a scaling plan you want to look at if you're using higher bandwidth trunks. That is, if you have potentially a smaller number of high bandwidth trunks and a smaller number of very high capacity routers in your core. Both these things help you quite a bit. In particular, if the number of routers that have to be concentrated in the core of the network is small, it makes some of the routing problems simpler to deal with.

When you reach a certain scale point, you also need to introduce a hierarchy to reduce the number of routers that have to exchange routing information with each other. A hierarchy might look like a central core with very high capacity routers, very fast switches and very high capacity trunks. Built around each of those core devices, there could be a whole regional network. An end user's traffic might start out in one regional network, enter the core network at some site, go across the core network and then go out into another regional network at the destination. In other words, there are feeder lines from the regional networks into the high speed backbone.

Junipers at OC48 and QoS Strategies

COOK Report: Let's return to your vBNS OC48 traffic map and talk about what the OC48 links will look like on the vBNS.

Wilder: We're in a new phase of testing regarding how we're going to handle OC48 trunks. We've had the Juniper routers in the lab since July 1998 and we've been happy with how they performed with workstations in the lab and with test traffic generators and such equipment. Now we're putting them out in a real world environment and letting them carry real world customer traffic. The routers are confirming their performance and doing quite well. Hopefully, before long this will be a jumping off point to designing a topology of the OC48 trunks around the network. This will allow us to put out our first traffic engineering model with MPLS label switched paths rather than the ATM core.

There are different QoS strategies that go along with this as well, like when we try to do something that's more scalable than putting reservations through the backbone for all of the reserved flows we need. Let me try to give you a high level description of this which is really another example of the usage of RSVP, i.e introducing the idea of hierarchy to QoS (in addition to RSVP usage in routing). The idea here is to have a very high capacity core in the middle of the network which doesn't have to know the details of each flow that needs a reservation or assured performance from the network. What the core does need to know is 1) that there are a few different traffic classes and 2) which packets belong to which class. For example. Suppose we have an assured delivery traffic class and a best effort traffic class. In order to reduce the complexity in the very high speed backbone devices, the devices would only have to look at a header field for each packet to see if it's assured or best effort delivery. They could then make a queuing decision based on that. In short, for now there are only two classes at the highest level of the hierarchy because that's probably the first thing we're going to do in the testing process. More classes are certainly possible.

At the edge of the network where the feeder lines are located, there might be less traffic aggregation. And as we said before, the less aggregation you have, the more unpredictable the traffic can be. Basically it means you have to be more careful with your QoS algorithms in these situations.

COOK Report: When we look at Diagram 1 (See Figure 1http://www.vbns.net/OC48/oc48traffic.htm) where you show Cisco in San Francico with an OC3c/ATM link into the FORE AX-1000, we also see a traffic generator. Is the joining point of the traffic generator and the AX-1000 analogous to where you'd have to transition from the mid-level hierarchy to the backbone one?

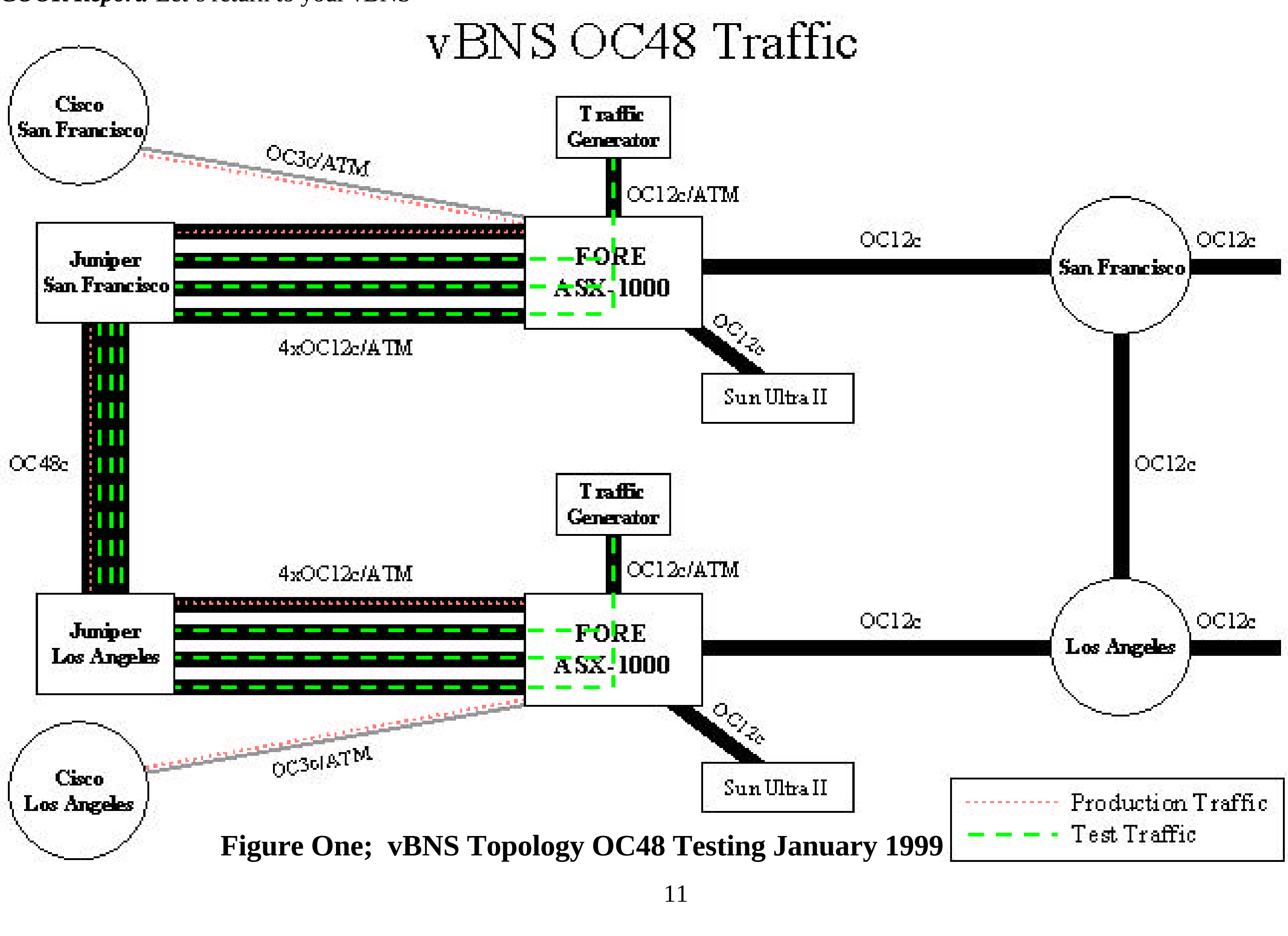

Figure One; vBNS Topology OC48 Testing January 1999

Wilder: Yes, it's analogous. But what you might actually have is a couple of dozen feeder lines (DS3s or OC3s) coming into Cisco from customer sites. *Those* would be the paths of traffic flow where you'd have to be very careful to reserve bandwidth. On those particular links we'd probably use something like RSVP for the critical flows.

As far as the link between the traffic generator and the ASX-1000 is concerned, remember that Diagram 1 only represents the trial configuration we've set up. It's not an architecture that we're going to use as we go forward. The object of the test depicted here is really to load up the OC48 test trunk between the two Juniper routers (one in San Francisco, the other in Los Angeles) and see if the routers can handle the load.

How do we fill up the OC48c link? The OC48c bandwidth is *extremely* high. And even though the research and education sites have plenty of high bandwidth applications, what they're doing at the moment can't fill the pipe. In addition, the highest speed of the trunks feeding this particular link is OC12. We need to have several times that load to fill up the trunk and test its ability to handle overload conditions. So we've added a traffic generator which is simply a device to spit out traffic. We use it to generate packets at OC12 rates to fill up the trunk. Then we create a point-to-multi-point circuit in the ASX-1000 switch. What this does is increase the input to the Juniper router. For every packet the test generator produces, three packets arrive at the router. We start out with an OC12 traffic stream from the tester and end up with three OC12 streams simultaneously entering the Juniper. So that takes up 75% of the OC48 capacity capacity right there.

As we go forward, the model we might use in the real world is that we'd have some number, say a dozen, OC12 customer interfaces coming into the FORE ASX-1000 and directly from there to the Juniper.

COOK Report: It would be helpful at this point to have some detailed information about the Juniper router, its architecture, and the engineering and design approach taken by the company.

Juniper Design Philosophy

Wilder: The box Juniper has built is based on the idea that you do all of the performance critical stuff [functions] in the hardware to the greatest degree possible. This enables you to handle the traffic at line rate and not be concerned about whether or not you're going to be able to handle it based on the details of what *kind* of traffic is coming in. For instance. Juniper has tried to avoid caching information about flows or routes- anything that would produce a situation where the cache would be flushed because of some change, like a routing change or a change in the actual traffic. If the cache is suddenly flushed, the router is effectively gone until the cache is rebuilt. Obviously, this is something all vendors are trying to avoid and it's a pretty high priority tenant of all the designs happening at Juniper. They don't want to be subject to this kind of thing so they build for the worst case scenario and don't have to make any assumptions about the amount of routing traffic or the amount of traffic entering. For instance. I don't think they're making very many assumptions about packet size. They know thay can handle a lot of small packets.

COOK Report: Why specifically would someone be concerned about making assumptions about packet size? How does the size of packets influence the design of routers and/or network functionality?

Wilder: HIPPI technologies are probably the best example where packet size is based on supercomputer characteristics. A Cray, for example, can push huge amounts of traffic but only in very big chunks or very large packets. It uses a 64 K byte packet or bigger and you can achieve a very high bandwidth with that. You can't do very many packets per second so if you're going to use small packets coming out of that Cray, you're not going to be able to generate much traffic at all. You won't be able to move the large files a supercomputer needs. That's sort of the opposite of the approach Juniper is taking.

In the real world, there are probably a large number of packets of different sizes that come flying at you. Now you could choose to design a router based on what you see on the network today. You could count how many small packets you see in a row, what's the average packet size, etc. and build a router based on those traffic characteristics. But if the load changes next year because there's a new popular Internet application, then you may have made the wrong decision. The critical factor here is where do you put your design points in the router. For example, if I'm going to build an OC48 interface, do I build in enough processing power to fill up an OC48 with 40 byte packets and still be able to handle that many packets per second? Or do you figure you don't have to do that because on average the packets on the Internet are 300 bytes, and you could use a smaller processor or put more of the functionality in software rather than hardware and you'll still be okay?

COOK Report: Say a little more about the design points.

Wilder: In the design process you really have to look at the requirements. And one of the requirements here is the number of packets in a router that have to be able to be forwarded per second. If you want to be completely safe, you take the minimum size packet it can possibly be and calculate how many of those will fit on the full trunk. If you can successfully handle that many, then that's as many as you'll ever have to do.

Another design point occurs when you assume the average packet size is something bigger than that so you'll only build the router to go as fast as the average packet size. The benefit to building it based on the average sized packet rather than the smaller packet isn't that you get more throughput in a given period of time. The main benefit is that it's cheaper to build the interface. Maybe you'll have fewer gates in your chip design or a lower chip count.

COOK Report: Do you think that the individual gigabit/terabit router companies like Avici or Juniper will engineer each of these design points in a slightly different manner?

Wilder: Yes. You're likely to have different decisions made when you're dealing with independent designs made by different people with different backgrounds. So each product will come out a little differently. And the more variety there is, the better chances you have that at least one product will be appropriate for a given set of conditions on the network.

COOK Report: How is the Juniper router different than a Cisco 12,000?

Wilder: I won't go into much detail because I'll leave that to a Juniper spokesperson. But I can say that Juniper has put more of its functions into hardware so the capability is there to handle a full range of just about any kind of traffic conditions. And that's something that looks very good to us right now.

COOK Report: It's my understanding that Cisco interfaces trying to handle OC48 are really having a tough time. Is that right?

Wilder: I really don't want to comment on this because I don't have the latest information from Cisco and I haven't seen all of their test results. I can say that I've haven't seen the kind of tests we're running on the Junipers done on Cisco's 12,000. But Cisco has been in this game for a while and they're certainly working on new ways of doing things. They've gotten a lot more sophisticated in their hardware support and they're doing very well.

But remember that the history of the whole router market is basically routers viewed as minicomputers. You pretty much did everything in software. The further back in time you go, the more you're going to see limitations in the processing of routers for any vendor. And if you're a new company starting from scratch, you'd want to go as much as possible for the hardware processing side of things as quickly as possible. Furthermore, if you're jumping in at the level of an OC48 capable router, then you simply have no choice. You *have* to build sophisticated hardware.

COOK Report: Can you elaborate on the type of decisions that have to be made regarding the use of hardware? What factors (like the organization of the routing fabric inside the router) influence these decisions? And how much functionality really resides in software?

Wilder: Hardware and software are both important because the sheer capacity to push the packets fast enough has to be there. If you can't implement all of the routing protocols, if you don't do BGP, IOS, ISIS and a few other protocols to boot, then basically you've got a non-starter no matter how fast you can forward the packets. Especially for new vendors entering the market, it's much harder and takes longer to get a good solid software base with all of the right routing protocol implementations than it takes to build the high speed hardware. Furthermore, you have to prove to a big network operator that you'll be a reliable box and interact correctly with all of the other peers you'll be communicating with. And wherever you go, people will know how to deal with Cisco's IOS. That's what they feel comfortable with and you'd better feel comfortable with it, too.

Juniper Software

COOK Report: Since software issues are important, what is it about Juniper's software that you find attractive? Its compatibility with other software? Its functionality?

Wilder: Juniper's software is looking quite good at this point in time. One thing they've done is to focus on what a big capacity core router has to do and not try to do all of the software functions that any customer premise router or a provider's edge router would normally do—such as SNA, encapusalation of various protocols and IP, or a whole slew of things that Cisco routers are good at doing.

COOK Report: What's the price of these boxes or are they provided for the beta test at a price you just don't know yet?

Wilder: We've purchased equipment from Juniper in the past and the boxes we're now using on the west coast were also purchased from them. As far as the price goes, they're cost competitive with the Cisco 12,000.

COOK Report: You've mentioned that about a quarter of the total OC48c traffic in California is real production traffic. Given this percentage of online production traffic and given the fact that you've had the Juniper routers in the lab since last summer, what kind of things are you seeing now in the network that you didn't see earlier in the test or in the lab when the conditions may have been a little different? Are there any changes you're expecting as you increase the percentage of production traffic in the OC48c link?

Wilder: The production traffic could have different timing characteristics from what we had in the lab. Who knows what could have been different? Fortunately, we haven't noticed any particularly new problems come out of the trial and we've been pretty happy with the way things have gone. An we've been running customer traffic on the west coast for a little over a month now.

COOK Report: How long do you feel you have to have the routers up before your level of confidence is high enough to warrant your putting in orders to install them in the rest of the network?

Wilder: I think we're nearly at that point. But there's also the time schedule for delivery of our OC48 trunks. We can't get ahead of that.

COOK Report: You are running through a single router an OC48c trunk that is really a combination of four OC12s? Is it that the hardware/software interface is more cost effective because you can use a pair of OC48 routers instead of four OC12 routers?

Wilder: Yes, that's one way of looking at it. Another possible way we could use this equipment in a real world scenario, as opposed to this test configuration, is to have a Juniper router driving two or three OC48 trunks going out to the wide area backbone and then have some number of OC12 or OC48 connections to edge routers. Each edge router may have dozens of customer connections coming into it. You also have to realize that if you have fewer routers and larger trunks, you have a smaller number of flows, less equipment to buy, less things that can go wrong, etc. In addition, trunking is physically and operationally less expensive if you have fewer eyeballs and hands needed for network support.

But there are limitations to how many OC12 trunks you can run in parallel because if you run a large number of trunks, it adds to the routing complexity. So it's easier to look at a smaller number of high bandwidth trunks because when you do nationwide routing at the trunk level, you don't have to deal with hundreds of OC48c routers. You only have to route between maybe a dozen or no more than three dozen of these types of routers. BJust being able to keep up with growing traffic is the main problem.

Dense Wave Division Multiplexing

COOK Report: Could you talk a little about the role of dense wave division multiplexing (WDM) and how it enables you to use multiple lambdas to provide quality of service. During the remaining years of the vBNS project, do you anticipate that the topology and type of testing you're currently doing will undergo any changes regarding your ability to run different lambdas? And finally, what effect, if any, does all of this have on the issues we've been talking about?

Wilder: You're really asking about the strategy of using different wavelengths to provide QoS. First of all, I need to point out that the trunk this traffic rides on is already using WDM at the transmission level. The OC48c in Diagram 1 is actually a SONET path between San Francisco and Los Angeles. At the optical transmission level between those two cities we have 4 or 8 times OC192 that this OC48 gets mutiplexed onto. So our transmission network has been using WDM for several years now. It's in no way new to us but so far the WDM does not change the way the routers see the bandwidth. I has no impact on them..

Looking into the future, we might have a different kind of transmission infrastructure where some wavelengths have extremely fast restoration times during failure and others don't. This might result in an availability difference to the end users who are using one wavelength vs. another one. So we might have different service types for IP users that have different availability or reliability numbers associated with them. The high reliability traffic would go over the fast restoration wavelengths and the low priority traffic would go over the slower restoration wavelengths. This is one possible scenario. It's not the one we decided to use. (But we haven't decided NOT to do this, either!). Since it's mentioned in some of the drafts on the optical Internet so I'm just throwing it out as an example.

COOK Report: Just because we now have all of these colors of light to play around with lower down in the stack at the transmission level, isn't it a mistake to think that you can also use these colors at a higher layer to do routing?

QoS and ISIS Protocols

Wilder: Yes. And the missing piece right now is that IP routing doesn't really do QoS differentiation. In other words, the ISIS protocol used to discover the paths through the backbone to find what it believes is the best path from point A to point B has a critical limitation. What it can't do yet is say: There are three paths between A and B. The first path is the best one for the premium traffic class. The second path is best for the low-cost traffic class. And the third path is best for some other traffic class.

By the way, the ISIS (Intermediate System to Intermediate System) routing protocol is probably the protocol where the most QoS routing work is now being done in the IP world at the IETF. That's because its being used by most of the very large ISPs for use in their backbones today. The IETF working group chaired by Tony Lee is doing a lot of work in the area of adding QoS differentiation to the routing. If you want to learn more about this, he's the one to contact.

The major purpose of the ISIS protocol work is to enable companies like Juniper and other router vendors to deal with the increasing need for QoS aggregation in the backbone and the desire to have m ultiple levels of quality in the backbone. Right now, all the differentiation of the different wavelengths is invisible to the router. IP doesn't know about it. What we're trying to add to the routing protocols is knowledge about the differences in layer one and the different capabilities of the underlying transmission network. That's exactly the goal here, i.e. to make the router aware of the service distinctions of the different paths so it can map the right traffic onto the right path. That way, when the customer marks a certain traffic stream as high reliability, assured delivery or some other characteristic, the routing protocol can match up the characteristics of the links through the backbone with the characteristics the customer requested.

COOK Report: What you've just been describing about how, where and under what conditions QoS is implemented sounds roughly analogous to the use of ATM in the core of the backbone when at some point the ATM VCs have to mesh with and interface with IP level 3 routing.

Wilder: I think it's more of an issue of how much knowledge the router has to have. Right now a router might see 100 virtual circuits and it probably feeds them just like it feeds any other point-to-point link, i.e. as if they were dedicated trunks. All it really knows about them using today's IP routing protocols is what the IP address is at the other end and some metric like how far away it is or how expensive it is. The router doesn't know things like, Does the link have a loss rate? Does it have a probablity of failure? Does it have a lot of delay or jitter? The router doesn't have any knowledge of those characteristics.

As a brief technical footnote, when engineers use the term "state" information about the network, they're usually referring to things that are *changing from moment to moment.* They aren't necessarily referring to the type of information above which is best described as categories of path characteristics. Path characteristics tend to be static things. For instance. One path through the backbone might be a satellite link so it has a very different delay than a terrestrial link.

COOK Report: Is Moore's Law and the increasing complexity of chip design having an impact on what we're talking about here?

Wilder: I don't think so. The more capable chip technologies and so on are really related to how fast you can do things and how much change in the network you can deal with from moment to moment, and not so much related to just having the intelligence to know a low-loss path from a premium path. As far as an impact is concerned, one of the trends I'm seeing is the industry moving toward standards that describe more things in firmware so that the routing fabric has to look at the packets as infrequently as possible. This area is really a distributed computing problem, i.e. making the routing protocols more complex. This is probably just as hard as making it possible to handle high capacity traffic but it's a different kind of complexity. It's the distributed algorithms and the routing protocols that people have to agree to, implement and debug.

COOK Report: How much are you learning on this project that will be useful to Qwest, Level 3, and other companies owning a lot of dark fiber with the potential of sending TCP/IP directly over glass? And what's happening within the vBNS that has relevance to the area of the convergence between voice and data networks?

Wilder: A lot of that has to do with what the end users of the network want to do and what advanced applications people actually try out. In one sense the vBNS is a public experiment because, as you know, it's a National Science Foundation (NSF) project whose mission is twofold: to do network research and to support highspeed applications. We do some of both. But as we evolve, we'll probably put more emphasis on the applications. Watching all of the research and education networks (CANet would certainly be a good example, as well as the vBNS and the other federal networks like NASA Science Internet and the ESnet) has been very useful to people who are trying to get a look into the future. Why? Because rather than experimenting with today's web applications, file transfers and telnet, etc., we're playing around with distributed simulation, virtual reality caves and high bandwidth multicast distribution of video. Maybe this is a glimpse into the future of what people want to do on the Internet.

COOK Report: Assuming Qwest or Level 3 could meet certain requirements, would they be able to purchase an industrial connection to the vBNS?

Wilder: At this point, they need to go to the NSF to get recognized as a player in this community. Then they could potentially get connected.

COOK Report: Can you say a few words about multicasting in the context of the vBNS and how you see it developing commercially?

Multicast

Wilder: There's a lot of interest in multicasting and it's a technology that's fun to use. However, there's no clear business model out there in the marketplace. The vBNS has been running native IP multicast for well over two years while the whole world was running the MBONE tunnels as the only available multicast service. We find that multicasting is a very interesting area even though it's pre-commercial in the sense that there's not a very clear business model for multicast services. It's also an area where the technology isn't very stable. If we were running a very large scale public Internet backbone, we probably wouldn't want to run high bandwidth IP multicast on the same routers with our unicast traffic forwarding. The software's just not as stable and the performance characteristics aren't as well known. The reason for this is that people aren't trying very hard to develop protocols because businesses aren't pushing them to do it. So this makes it an interesting area for networks like the vBNS which allow people to play with it and see what applications can be made to work well on it. Then maybe it will become clear as to what the business model should look like, what some of the reasons are to commercialize it and how to make it ready for the public Internet.

One of the problems in the development of a business model is that when the traffic crosses network boundaries between providers—say UUNET and BBN—there's no agreed upon method of accounting and/or settlement for the shared resources. Another problem is just that if you give any user the ability to send out data, you won't just be loading one path through the network. You may be loading 100 paths. From a commercial point of view, you'd want to charge more if more paths are going to be loaded.

COOK Report: Is there any way to handle a situation like that? Can a user buy the authentication that would allow them to multicast up to a predetermined amount? How would the accounting be handled in this case?

Wilder: There are no definitive answers to these questions yet. And that's part of the reason why commercial development hasn't taken hold. With all of these critical issues unresolved, there's no quick payoff for a business. But I think something commercially viable will eventually emerge from multicasting. And what will help the process along is allowing more time for people to play around with it and decide what's really useful. This is what will motivate people to come up with a way to make a viable commercial application. Right now it's a bit of the chicken and egg problem. If there's not a clear demand from the commercial network providers because they don't see a business model, they're not going to push the vendors to do the implementation in a serious way and harden it for their use. So it's never going to get started. We're trying to solve the chicken and egg problem by making multicasting available to a small community in a safe environment where, if they use a little more bandwidth or if something fails every now and then, it doesn't have quite the same severe consequences it would have in the large scale public Internet.

Some of the other things we're doing regarding multicasting is that the vBNS is also carrying the Internet 2 digital video. The distribution community can have up to 6-7 channels of high quality video. Each channel can be received by multiple people who subscribe to multicast. Each receiver sees about 6MB per video stream.

IPv6

COOK Report: Does IPv6 perform more of an administrative task? It doesn't have to be on a vBNS to be useful, does it? Or are there certain things that would make IPv6 more congenial to a high bandwidth backbone network than the ordinary regional networks?

Wilder: This is another area we find very interesting because it's one of those technologies that isn't a real commercial technology yet. Maybe it will become one if people are allowed to play around with it enough and do sufficient research and development. It's the chicken and egg problem again. There isn't enough IPv6 infrastructure out there to convince anyone to slow down enough to experiment with adopting it.

COOK Report: Has the development of NGI (Next Generation Internet), Internet2 and Abilene affected your research and education usage?

Wilder: Abilene is just starting to come up now. We have peering with them. We're willing to play in this environment as one of the R&E backbones together with Abilene and anybody else who comes along.

COOK Report: Tell us about the organizational state of the vBNS, where it fits into the research picture and more about the users. We understand that you have pretty stable core community of university and research oriented institutions and that all of their connections are working smoothly.

Wilder: Yes, things are going quite well. We've seen nothing but growth so far and anticipate that it will continue. (That's why we're looking at the OC48 capabilities.) We have over 80 universities connected as well as supercomputer centers and several peerings with other federal networks and international R&E networks. We're willing to go forward into this multi-backbone model if that's what the NSF wants us to do.

COOK Report: What are some of your goals for the next few years?

QoS Futures

Wilder: QoS is very high on our list of goals. We want to be able to do both the end-to-end reservation model and a more scalable class of service model together on our backbone. In fact, we're working with Internet2 on this very thing. They have an activity called Q-Bone that we're participating in to try to do QoS between multiple providers. This is important to us. We want to interface with other backbone providers in order to have multi provider QoS and related services to provide to our users.

Progress in this area is slow because the issues are very complex. There's probably some feasible technology that can be deployed and some feasible solutions, but we need to have enough awareness of them and build consensus around them so that we have a reasonable approach to go forward with. Talking about multi provider QoS strategies doesn't make sense unless there's a pretty good consensus about where you're headed. Otherwise, the multiple providers can't implement equivalent services.

Let me close by saying that QoS in a connectionless Internet packet network *can* be achieved. It's just that conceptualizing the problem so far has been a daunting task. And for very good reasons. Breaking it down into its component pieces and getting enough different efforts working together to build the whole architectural solution for a problem as complex as this one doesn't come quick and easy.

COOK Report: Are you more optimistic now about achieving this kind of QoS than you were 3-6 months ago? Is it more a matter of economics rather than simply IP protocols?

Wilder: That's certainly a big part of it. And that's why we're trying to roll out very soon a Dif-Serv on net within the vBNS QoS offering using this reserved bandwidth idea. Even though it's a chicken and egg problem, we're saying that we should go ahead a make something available so that people can play with it to see what works and what doesn't. If we do that successfully, it won't do anything but fuel the discussion and standards work dealing with the issue of how we'll do a "standardized multi provider solution" a little farther down the road.

COOK Report: Let's discuss the sequence of events that will make this happen. From a standards point of view, do you feel that there's now a conceptual understanding of the total number of pieces that have to be in place to produce something worthwhile? If this is true, when you get further down the road and see the technical solutions coming into focus, will that provide enough motivation to begin seriously looking for a way to handle the cross-provider settlements, the economics and the political issues underlying all of this?

Wilder: Yes, that's the sequence of events that has to happen. Without having the applications and the end users demanding the service, there isn't gong to be any motivation to move forward with the economic and political solutions. We're just trying to provide the environment where people can run the applications and see what's valuable. Hopefully , they'll then get fired up about doing something specific in a serious way.

Internet Telephony and QoS

COOK Report: Are any of your users demanding Internet telephony?

Wilder: Internet telephony is certainly going to be one of the drivers. I'm not really feeling a lot of pressure today from people to do voice over the Internet but because I work for a telecommunications provider, there's certainly a lot of thinking about it. It does fit into some of our plans, however, because it's an excellent test application for QoS mechanisms. If you start deploying a reserved bandwidth service or a Dif-Serv offering of any kind, running voice end-to-end over the service while you have best effort traffic saturating the bottleneck points in your paths is probably the most convenient way to test how well it works. You can usually hear the delays, jitter and loss as they occur in our end-to-end connection.

COOK Report: Some people say that QoS

Continued on page 24

Two Views of Electronic Commerce

Eyeballs Versus the Cluetrain Manifesto
TV with a "Buy" Button or Self Organizing Markets?

Editor's Note: While *The COOK Report* is by no means ready to abandon its focus on internet infrastructure (engine rooms and the underlying transport technologies), we have been giving some thought to taking occasional looks at seismic movements taking place among the higher levels of the protocol stack. Having read the Cluetrain manifesto (see below), we wondered how to bridge the gap between its apocalyptic views and the current stock market inebriation with anything having the odor of Internet commerce. Thinking that perhaps an expedition in the direction of electronic commerce might be in order, we have taken advantage of an opportunity both to review a study on e-commerce "Portals To Profit: E-commerce Business Models and Enabling Technologies"(published 4/16/99 and available from www.datacommresearch.com) and to interview Chris Locke (one of the authors of the Cluetrain manifesto) on March 28.

Part One: Portals to Profit

E-commerce studies by IDG, Gartner, Jupiter and others abound. Our ability to review this is handicapped by the fact that we are uncertain about how closely the quality of this study (whose primary author is a venture capitalist) stacks up with the quality of other such reports. Therefore we will present the general conclusions of the e-commerce study and comment on our own perception of its insightfulness. The report is by Michael Hentschel of Techvest International (a venture capital firm) and Ira Brodsky of the Datacomm Research Company. According to Brodsky: "Our main thesis is that e-commerce competition will transform all commerce, destroying many if not most traditional business models, and forcing companies to invent new ways to make money. Thus, it is imperative that businesses understand the new models, many of which involve cost-based pricing, below-cost pricing, auctions, reverse auctions, ad-targeting, etc."

So far so good. But the report goes on to paint its picture through the eyes of the Fortune 500. E-commerce is primarily about size and, as such, is a race to see how the large corporations can translate their standard views of the world into this new medium. The views are basically the old industrial age mechanics of economies of scale using, this time, the latest digital technologies and suites of software agents behind the scenes to manipulate and shape the thoughts and response of the customers. The customers are portrayed mechanistically as sets of eyeballs on which the new technologies act. The object appears to be to get these eyeballs thinking they are still quite independent when in reality computer enhanced versions of standard marketing techniques are being applied to create the highest probability that they will snap up the "bargains" placed in front of them. For someone who has been on-line for nearly 20 years, framing it this way leads us to wonder whether those who propose to use the Internet in ways that merely speed up the standard approaches of marketing and selling understand adequately what they are dealing with. While we are less than fully comfortable with the report's point of view, it is far too early in the development of electronic commerce to say that it may not turn out to be correct in its key predictions.

What the Study Says

Let's review its basic points. It finds that . "Savvy use of information technology (IT) will be key to achieving competitive advantage for at least the next few years. . . [and that]. . Opportunities abound for at least the next few Internet generations: (. . . .) Those prepared to manage end-to-end supply chains will lead the way.

But there are also many dangers: (1) The large number of players chasing each opportunity will help to create some very big losers; (2) Market capitalizations will not remain as high and forgiving as at present; (3) Business models based purely on future advertising are in many cases "recession-bait"; and (4) "Virtual business" models will often lead only to virtual profits." (p. 7)

"The following are some of the key trends; technology is critical to each: (1) Portals will remain the usual starting point [of e-commerce gateways]. (2) Hubs will help guide users [further] on topical subjects. (3) Commerce destinations will be sorted by price agents [designed to find the cheapest prices]. (4) Pricing will become a less useful tool for making profits. (5). Search engines will perform meta-searches using multiple engines. (6). Personalization data will guide search agents to the sites that matter. (7). Customers will be able to quickly and fully educate themselves about any product or service. (8). User choices will trigger real-time transactions and logistics software (i.e., flow of goods from factory to warehouse)." **Editor's Note**: Point seven may well become true, but if it does, it would seem to lessen the importance of advertising which plays a role of critical importance in this business model.' (p. 15)

The study then moves on to discuss advertising pointing out that:

"Alliances + Eyeballs + Technology = Ad Revenues
* E-commerce combines goods and information flow.
* E-chain alliances control users, delivery, and content.
* Leaders are driven to form conglomerates around new technology standards.
* The e-commerce experience captures audiences in any electronic media.
* New brands are formed with tremendous profit potential.
* Ad targeting and ad revenue control become the grail of profits. (p. 16)

"Since the magnet for ad revenue is the number and quality of eyeballs that arrive at the site, information, and even software, can be offered free to the user in exchange for their more scarce attention span. If ad revenue is sufficient for leading players to more than cover the declining costs of gathering and distributing information, i.e., profitable, then all other players will be forced to embrace the same basic model." (pp. 20-21)

"E-commercialization is the complete transformation of commerce for goods and services into valuable (and therefore chargeable) information streams. Charges take the form of cash payments for goods and services, and cash or non-cash payments for accepting advertising messages. Increasingly, information streams will possess their own attached, targeted commercials, paying the bulk of the cost of both the information and its delivery. Thus, even the physical goods become a loss leader (see Buy.com) in the effort to create the size and quality audience necessary to attract advertising." (p. 22) In other words because the technology makes it relatively easy for users to establish who sells equivalent goods for the lowest price, price discounting will become a dominant approach. Whatever e-commerce site has the biggest advertising revenues will be able to offer the largest discounts and still remain profitable.

The study lists 12 major conclusions - several of which follow and most of which we find reasonable.

"1. The Internet is the world's biggest, fastest, and most accurate free market. It is rapidly becoming the most efficient sales channel ever devised. Any business that faces competition will have to participate in Internet-based e-commerce. Most previous e-commerce forecasts will prove low.

2. Atomization (creative destruction) will transform all commerce into electronic information chains. Old business models will be destroyed — but with a purpose. The old models will be broken into pieces and rearranged on the Internet and extranets in every conceivable manner. This great transformation will accelerate as IT budgets are freed from Y2K allocations.

3. The best-capitalized portals and hubs will pull well ahead of the pack, using their stock market valuations to acquire whatever technology they need. However, there will still be opportunities for "e-tailers" who add value by helping customers find what they want or need. Thus, money and knowledge will become interchangeable on the Internet.

4. Mergers of big portals, ISPs, and telecomm access providers will abound. Through such mergers, portals will strive to emulate AOL's subscription model. High-speed access providers will acquire brand name portals, in an effort to deliver new forms of content to captive audiences. ISPs will be assimilated by big portals and major telecomm access providers. But there will be temporary opportunities for new foreign portals — opportunities that will gradually disappear as real-time translation software breaks down language barriers.

7. E-commerce will require new business models and much experimentation. Products sold at cost or seemingly given away for free will force businesses to invent new ways to make money. Some businesses will sell products at cost, making money off advertising, shipping and handling charges, membership fees, and even cash flow. Alliances will be crucial to gaining proximity to favorite destinations and sharing traffic flow. E-mail marketing will often replace direct mail. Auction sites must evolve, as intelligent agents will turn the entire Internet into a real-time auction.

9. Browsers have become commodities and, therefore, largely irrelevant to business differentiation. Metasearch agents will undermine existing search technologies used by the big portals. Personalization technology will benefit buyers, enabling personal portals, and sellers, enabling better ad targeting. Search software with intelligent agents will prove the long-term key to relevant data and relevant ads." (pp. 22-23)

The study goes on to talk of e-commerce business models find no fewer than 20 grouped into four categories.

"Conventional - Basic, Low-Margin, Community, Brokering
Competitive - Services, Price Agent, Zero-Margin, Sub-Zero, Free/ Sub-Free
Niche - Entertainment, High-Margin, Vertical Hub, Auction, Virtual World
Relational - Personal Portal, Toll-Taking, Biz-to-Biz, Keiretsu, Super Agent, Nirvana"

Hedging One's Bets

The study continues: "It seems unlikely that more traditional models can survive in the face of the success of the more aggressive models. Yet the economic theory behind the most aggressive strategies are in some cases suspect, and it is possible that more conventional strategies using good marketing techniques will prove better long-term survivors." (p. 25) The study uses a brief discussion of each business model to introduce readers to the complexity of e-commerce web sites and the technology surrounding them. It concludes: So far, only big corporations have implemented large-scale, strategic applications on the Internet, positioning them way ahead of smaller enterprises. However, Internet applications are highly-scaleable: they can not only help large organizations extend their reach, they can put sophisticated capabilities in the hands of smaller players." (p. 33)

One of the most sober paragraphs in the study admits: "We have inquired how much more advertising spending would have to occur to replace all the profits the new business models are foregoing on the assumption that eyeball-based advertising will more than make up the shortfall. If a quarter of the projected 1999 Internet economy (about $25 billion) were to retain current profit margins (not all that large), and half of the Internet economy were to adopt a lower gross-margin model as a result of price competition, and another quarter were to adopt a zero- and sub-zero margin model, then we are talking about offsetting $10 billion in combined losses. Even if we assume Internet productivity will rise, this is more than 2 times the likely $5 billion in expected 1999 Internet advertising revenue — after a record $2 billion in 1998." (p. 49)

"Once equity capital stops viewing Internet losses as positive "investment," and more aggressive business models create further gross-margin and profit-margin pressure, reality will set in. Real profits must be found somewhere — and relatively soon. The great hope is that advertising revenues will increase dramatically." (51)

"Financial markets are valuing the total electronic economy at a high multiple of currently available dollar transactions — in other words, sales — with little regard to profits or even future profits. Where once the standard for high-growth companies was about 30x EPS (assuming a 30% growth rate), it is now 10x-30x or more of sales. Assuming 100%+ annual growth rates and, therefore, 100x EPS valuations and a 10% after-tax profit stream over a very long period of time, 10x sales almost makes sense. Small details get in the way, however: multiples of sales are now even higher than 10x; 100% revenue growth rates are not sustainable for long; and after-tax profits are nowhere near 10% over the foreseeable future. Is the stock market irrational?" (p. 54) The study goes on to build a case as to why the market may not be irrational.

The study concludes with a long section describing the importance of a long list of largely software technologies that are use to construct large Internet commerce web sites. These sites, if they are fully integrated with all the information systems of a large corporation, can cost over 5 million dollars a site and an equally amount to run per year. It offers predictions of various Internet companies and strategies that will fall both into the likely winner and looser categories.

The study is certainly a good introduction for the non expert. It covers a lot of ground but does so often with a level of superficiality that we found frustrating. For example on page 71, there is a short section on middleware that says: "the scalability of Web applications often hangs on the inability to upgrade and maintain applications on users' computers. New applications are licensed annually, administered externally, and preferably outsourced entirely. This is taking hold especially in the Internet arena, where everything is new, and the expertise seldom exists in-house to implement the latest advance. Here is where middleware comes in, almost invisibly supporting developers inside and outside the Web site." (p. 71) The study then lists ten middleware vendors and moves on to a section called "Database Integration technology for the Enterprise and the Web." We would have found a couple of paragraphs explaining what exactly at least one middleware application is to be helpful. Of course the authors may argue that had they done this throughout the entire study it would have been twice the length and never could have been sufficiently up to date to publish — given the fast changing nature of the field. We must also acknowledge that we like a level of detail that may go well beyond the requirements of ordinary readers.

For anyone at a major corporation trying to get up to speed on a subject that undeniably is reshaping the national and world economy, the report is well worth the purchase price. For those who already have expertise in the subject, the authors assure us that the report's scope and business model analysis make it unique.

As Ira Brodsky informed us: "Portals to Profit is a report for business planners. It is very aware of the services side of the Internet, including the potential to automate customer service and, in many cases, deliver new levels of customer satisfaction; [and] it recognizes new opportunities for small and niche players. However, the Internet will also create new corporate giants, and some old giants will eventually figure out how to succeed in the new environment. To wit, it's amazing what you can do with a few billion dollars."

Undeniably the focus of the report is on the muscle power of the largest players. Consider for example the following conclusion: "The bulk of ad revenues will continue to go to the top 50 sites. Ads allow the top portal sites to offer search services and information for free, creating tremendous competitive pressure on smaller sites that cannot attract significant ad revenues." (p.75)

It certainly represents a legitimate point of view and its focus on automated mechanistic ways of dealing with the masses describe a modus operandi that may be all that is needed for success. After all 90% of the people on the net have been there less than 2 years. They may never be able to see the Internet as much else besides TV with a buy button. However another view exists. The Cluetrain manifesto suggests that the Internet makes a whole series of special relationships between customer and businesses possible. It is the view of Chris Locke. Someone who has, as have we, been on the net for nearly 20 years. We believe that, the longer one is on the net, the more the experience broadens one's horizons.

In this world view the web is driven by the corporation's customers who can join together to route around companies that can't or wont meet their needs. Web enabled consumers, according to Cluetrain, are there not to be manipulated but to join together in efforts that will replace those companies that: "don't get it." Cluetrain represents both a culture and ways of looking at Internet commerce that business would be very much ill-advised to ignore.

Two years ago the trade press was harping on the alleged fact that a shake out for small ISPs was inevitable. Consolidation into giant national ISPs that would provide economies of scale would happen overnight — given the hundreds of billions of dollars being invested by the major telcos. Indeed much consolidation has taken place but, contrary to the predictions of the experts, there are many more ISPs in business now than there were when the consolidation started. The growth of the Internet and its ability to create and absorb new applications is one reason why. The fact that with low operational costs and well focused market niches "mice" like *The COOK Report* can continue to prosper underneath the feet of elephants is another reason why. It is this point of view that Locke (being himself another "mouse") is well aware of. The authors of Portals to Profit are also aware of it, but given their target audience, it is understandable that they emphasize it much less. Our readers will be well served if they can assimilate both points of view. Figuring out the Internet and e-commerce is a complex task on which no one can yet claim a monopoly.

[**Editor's Disclaimer**: Ira Brodsky is reselling our IP Insurgency study.]

Part 2: The Cluetrain Interview

Chris Locke crashed our radar screen 1993. In 1994 we reported on some of his earlier iconoclastic activities with MecklerWeb. He had designed an alliance of companies that would function as a portal. The only problem is that back then no one had figured out that there might be a market for such a thing. His latest product is the Cluetrain Manifesto — www.cluetrain.com — a powerful screed that he and three fellow instigators have just nailed to the doors of corporate America. We have here 95 theses that purport to explain what most large corporations do not yet understand about the Internet. The Internet is the catalyst which sparked the Cluetrain manifesto. The Internet is also a medium that promotes styles of communication that are deeply subversive of standard ways of doing business. The Manifesto shows why the web empowered voices of millions of small businesses and consumers may overturn the industrial age control-oriented business models of giant companies.

Five years ago, with Mecklerweb, Locke saw the web as a tool by which markets could be radically restructured to bring companies much closer to their customers. He imagined that, once the corporate decision makers were shown the ways in which the web could open all kinds of new communication channels, both between the company and its customers and within the corporation itself, management would share his vision. In short he was certain that management would become just as enthusiastic about using the web to reform the way the corporation did business as he was. He was wrong.

That was five years ago. Since then the web has swept the Internet. Everything and everyone is on the world wide web. "I web there I am." After Alan Meckler pulled the plug on Mecklerweb in September 1994 Chris did tours of duty at MCI and IBM. There Locke found that corporate executives saw the web not as a transformative tool but as a mirror, the purpose of which was to reflect their own views of an orderly world amenable to the continued top-down controls of the industrial age. While Locke saw pockets of innovation in the large corporations, control-insistent hierarchies usually conquered the efforts of the innovators whose temperament he shared. The Cluetrain Manifesto marks the "revenge" of the innovators.

Since the internet served as a low cost platform for further experimentation and development of the vision Locke had seen in 94, he launched his own business to teach those who would listen how web sites could be used creatively in the internet medium. At a 1996 Esther Dyson retreat that gathered industry creative types together to bounce ideas off each other, he gave a tub thumping talk on how the big folks just didn't get it. "Give'em hell rage boy," shouted someone from the back of the audience.

The moment struck a chord and Locke adopted the online personna of RageBoy. Why not use a web site and a mail list to proselytize for what he believed in? He launched his own mail list or e-zine called Entropy Gradient Reversals (www.rageboy.com) under the motto "all noise, all the time." With his focus on what the web could be, he began to publish a regular series of "rants" where he satirized those who didn't understand the new paradigm and praised those who did. After nearly 3 years with 3000 subscribers Locke has the ultimate Internet marketing vehicle - a means of acquainting potential web customers with his views. Having been an EGR subscriber is a requirement for sitting down with Chris to discuss buying his other services.

The web-based marketplace has thrown Chris together with many other iconoclasts. Cluetrain is the exposition of where the loose fast and out of control web and internet is headed. It is the product of Locke and three other co-conspirators. According to **Locke:** "The four of us: David Weinberger, myself, Doc Searles, Rick Levine probably have somewhere on the order of 75 to 80 years combined experience in being online. We also have considerable experience in doing various forms of marketing and Web stuff and interacting with business. Consequently what we have formulated in the Cluetrain manifesto is not very theoretical. We are boiling down the experiences that we have heard from many people and that we have had ourselves. Much of the response we're getting is along the lines of: why did this take so long? We know this but we also believe that it is the first time that it is been all articulated as strongly and in a single place."

The Internet Is Not TV With a Buy Button

COOK Report: What exactly have you done?

Locke: We have created a synthesis of ideas, many of which, have been in the air for a while now. I also see this is a kind of gestalt that might inform your perspective and that of your readers over time. While right now you are focusing mostly on infrastructure, we see the Cluetrain manifesto as more consumer oriented. Cluetrain, at its most basic, is emphasizing how different the Internet is from broadcast.

COOK Report: And a lot of people are still having trouble realizing that?

Locke: The people who have trouble realizing these are people, whether not they are aware of it, who, in their heart of hearts, have trouble realizing that the Internet is not just TV with a buy button. These are executives are making decisions to allocate hundreds millions dollars on Internet and Web applications but who have no real first-hand experience of the medium.

COOK Report: Nothing ever changes. Dave Hughes in 1980 and in 1981 went to Washington as the largest single customer of the Source to talk to Source executives about what he liked and did not like and found that these executives never used their own systems either.

Locke: I'm not terribly surprised. New technology comes down to people who tend to evaluate it only in terms of technologies that they know. Thus when photography appeared, people went: oh, that's sort of like painting. And then movies came, and people say: oh, this is like the stage. We'll make a movie by bolting the camera to the center front of the stage where it can capture the action. And then television comes around, and people say: hold on. This is just like motion pictures. McLuhan said it: all new communication technologies are initially perceived in terms of their predecessors.

The Cluetrain manifesto is about the vast difference. The Internet is perceived as television by people who haven't used the internet. You will not understand why the net-as-TV inference is wrong unless you have spent a lot of time on the net itself.

COOK Report: But isn't Cluetrain derived from a broader foundation than that of the mistaken harping of marketing executives? Doesn't it reflect expectations from a growing number of people who are spending more and more time in the online medium as to what can and should be done with the medium itself? But the reactions of readers have been powerful enough so that it seems to go beyond just this. What are you saying that provides this extra spark?

Markets Unconstrained by Hierarchy Are Self-Organizing

Locke: One thing we're saying that we think is quite powerful is that markets are learning and self-organizing at a faster pace than companies. Organizations won't even face their customers unless their information is coordinated first. But coordination takes time. Nevertheless, the big product here has been Lotus Notes. Unfortunately Lotus Notes is a top-down installation which requires that the IT Department do the systems analysis and spend weeks asking people what they really need in the way of forms and macros and things like that.

The open marketplace can organize without the constraints of hierarchy, of bureaucracy, and of command and control that saddle most organizations. If you take the curves of learning over time and graph them on a chart, you have the shape of a hockey stick which represents the marketplace. It's sort of flat for a while, but then when it starts rising, it rises like a hockey stick. In this open marketplace, ideas are tossed from person to person with great rapidity. Try this. No, this works better. Boom, boom, boom. It's like Linnux. We just had Eric Raymond, author of the Cathedral and the Bazaar and President of the Open Source initiative sign up with the following comment: "the Cluetrain is to marketing and communications what the open source movement is software development — anarchic, messy, rude, and vastly more powerful than the doomed bullshit that conventionally passes for wisdom." It is a territory where no ideas are off-limits. Whatever works goes.

If you map this, you have markets that are learning very very fast. While, at the same time, you have Corporations, with their emphasis on hierarchy and control, almost flat lining it. Right now the open market curves are under corporate ones, but they, are coming out very fast. Where the two cross will be the point at which everything changes. What we are saying is that the open market learning curve will be shaped like a hockey stick - rising very suddenly and steeply when it begins to learn. We are already beginning to experience it. It will be a discontinuity. Think chaos theory. The big corporations will literally not have time to react.

We are not crystal balling this one except to say that these companies could die as a result of this change. Why? Because this smart market — out there on the Internet is deconstructing the marketing messages coming out of these large corporations. One of the biggest uses of the Internet is to send around joke mail. You know — things like: Microsoft buys the Catholic Church. But you know it's not just Microsoft. It's everyone. Someone gets some clueless home page and suddenly you have thousands, hundreds of thousands of people are rolling on the floor and laughing and going oh my God look at that — I can't believe these bozos. You know how fast word like that travels.

But there are many big corporations that don't have their ear to this kind ground. They go to their corporate meetings; pat each other on the back; congratulate each other for being written up in all the right places. And all this happens while the marketplace is saying: yeah, we will buy your technology a while longer — so long as is not totally busted, but we're just not impressed that you understand what's going on.

Corporate Pachyderms Rendered Superfluous by the Decreasing Cost of Technology

COOK Report: Well if you look at the role of technology in the century just ending, you will see that at the beginning corporations were needed perhaps because the industrial age called for a size, scale and scope that only a giant organization could provide. The cost of entry into new markets and businesses was so high that only a large corporation could cope.

Locke: Yes if you wanted to put together the next car company or the next Hurst or Gannet publishing Co., the cost of entry was vast. But Web reduced the cost of entry to near zero.

COOK Report: Consequently as long as everyone can interconnect and communicate, they can self-organize.

Locke: And a really good example of this is MP 3. This compression standard came out of a place called the Froenhaufer Institute in Germany — your standard academic research lab. Within the last couple of years the technology has leaked out onto the net. The compressor, a very proprietary piece of software, was grabbed by some French teenager who, thinking it to be freeware, distributed it, worldwide. Now when that happened the authorities found out about it fairly quickly and came down upon him like a ton of bricks. Now once this happened, it was too late for a friend of mine who had it on his FTP site in Texas did not even know that he was in possession of anything of significance. Consequently, I was able to grab one from Texas, and another from Korea and so on.

If you take a track from a CD, you will find that a typical rock song is easily 50 or 60 MB. The MP 3 compression ratio is between 10 and 12 to one. As a result that 50 MB Nirvana song may be easily compressed to about 4 MB. And at four megabytes you can even email it to a friend. Now I found a Mariia Cary cut about a week after her album was released. You'd have to be an audio file to really be able to tell the difference if you were to do a blindfold test between a commercial CD and the MP3 playback. It is like light years beyond real audio. It does take some time to download but it has changed the economics of the 50-billion-dollar world-wide recording industry and has that industry quaking in its boots. The Recording Industry Association of America (RIAA) rattled its sabres and said that it had this Web "bot" that was going to go and find these sites offering the MP 3 files and shut them down. But at the same time I'm seeing on Usenet lists announcements of FTP sites run by these warez kids saying: here's a new Site but come quick it will only be up for the next eight hours. Here is the listed stuff you can download, but hit it quickly before we change the IP address.

This is smart markets in operation. Copyright is an obstacle. We can route around it. We are faster than you and we know more than you. I said to myself the other day: these kids are just goofing off. If they ever get serious - game over.

COOK Report: So, in the area of electronic commerce, rather than trying to replicate the industrial age you had better focus on the issue of how you plan for and cope with these kinds of technology changes?

Locke: Yes, and from this point of view, MP3 is again a good model, because Tom Petty has got an album coming out in mid April. And almost a month ago now he went to MP3.com and said: Here's one of the prime tracks from this thing. I will give it to you in MP3. Distribute it to the world for free. Now you know that this guy's album, when it comes out, will go to the top of the charts.

This is one of the differences — markets getting smarter and organizing faster than can the company, which thought that it owned the markets, but, in reality, doesn't know what its dealing with anymore. One of the things that we say about this — and its really important — is that smart markets are like the Internet itself without a centralized brain. Did you ever see a flock of birds flying formation very close to each other and suddenly changing direction without any of the birds colliding? There is no lead bird. They suddenly just all swoop — gracefully and rapidly changing course.

Markets Can Change Suppliers Overnight

Like the flock of birds suddenly changing direction, Markets can change their suppliers overnight. If it is perceived that this company is clueless; that they no longer have the vaguest notion of how to respond to us, the change will not happened gradually over time. It will happen between 3 and 5 p.m. on some random Thursday that they will realize that their market is suddenly gone. It will happen because someone else came along who understood the community, knew exactly what the requirements were and offered something that they could download easily. Now in fact there is a rather large example of this that actually did permit a couple of years of response time. The company involved is Amazon. But by the time Borders and Barnes and Noble took Amazon seriously, it was too late to recover.

COOK Report: One of the best uses left for their physical stores is as a place to go if you actually want to hold and look at the book before you return home and by it from Amazon.

Locke: The point is there are other sources of revenue for the physical stores, but if Amazon takes enough of their market share away, their margins are shot to hell. Now I saw this ad online in the Industry Standard the other day and it left me in stitches. It was an IBM ad for their e-commerce campaign. "IBM, your e-solution provider," it purrs. And then it goes: Borders — to whom did they turn when they wanted to come onto the Web? They turned to IBM.

I know the story behind this. I was on Salon — a conferencing system and a big big e-zine one that is very slick — the day that they came up. In fact they came up, and crashed, came up again. On the first day they said "sponsored by Borders." Now I'm a big book fan and I like Borders so I said: oh, cool and I went clicking over to find Borders to see what they had in their electronic store. All I found was: "coming soon." Now that was like late '95 I am going to say because it happened while I was still in Connecticut. I thought: oh well, they'll be up in a few weeks. It was a few years! I kept going back and saying to myself: what's taking you so long? Finally I talked to an EGR subscriber who worked for Borders who said: "don't ask. You don't want to know."

COOK Report: The corporate hierarchy was probably pissing and moaning about every little step.

Locke: It is not just that. Here's this ad from IBM saying: we are responsible for having brought Borders onto the Web, and I am going: you are responsible for having lost them their business because you took so long. You gave Amazon a two year head start over them you idiots. The point being that it to the unwashed, that IBM ad would look very persuasive. But to someone who really knows, it's like: you brought Borders online two years too late. What are you crowing about?

And who can forget how IBM screwed up the Olympic reporting a couple of years ago when they were in charge of all the new feeds to something like a 150 international news organizations? It wasn't just that their Web sites couldn't handle it. They also hosed the data. They had video feeds with people who looked eleven feet tall wrestling people who looked two feet tall. Things stayed messed up through most of the Olympic's and they had the international press just screaming. Now can you imagine? I went to their Web site while all this mess was happening and it said: "and we can do this for your company."

The point being — by contrast — Amazon started out as a little company "that knew." Amazon's competitors paid no attention to its Internet strategy, but, as we know, things changed really fast. And those people are really scared to death of Amazon today. In fact they're not just scared; they are bleeding from them. This is a story that will be repeated over and over.

The Freedom of the Web is the Freedom to Have Your Own Voice

Now the other aspect of Cluetrain that we think is definitely new is that there is a particular style of conversation on the net. That this style facilitates, enables, and mediates. This style is voice-to-voice. People talking the way you and I are right now. People are telling war stories. But the language that they are using this completely antithetical to that of the press release, the language of the annual report, and language of the dog and pony pitch of the big corporations.

COOK Report: Part of what you are talking about is whether or not there can be trust in these communications channels. And from your description there is precious little trust. Ed Gerck is a specialist on authentication who talks about trust as something that can be delivered only out side of the channels by which people are communicating at any given moment. That people will recognize a style of communication with which they feel comfortable. Does that make any sense in the current context?

Locke: Yeah, it does because we all bring a varied inventory of knowledge and expectations to our communications. For example because I know you, when you call me up, my reaction to your call will be very different than it would to a salesman's cold call.

COOK Report: I tend, in part, to base my decision on whether or not to open a message read it, on my knowledge of and trust in the author based on long past experience reading other postings by that person.

Locke: Absolutely. Here is another gating factor which is getting to be a finely honed skill. How many seconds does its take you to recognize a spam message? I have gotten it down to about a tenth of a second. The cues are very subtle. But given the increased volume I have developed a hair trigger relationship with my delete key. The same thing governs my relationship with the delete key and visiting new Web sites. Now there are companies out there reporting millions of its day, but if they were honest, they would report that the lengths of stay of a large number of their visitors on their WebSite is measured in seconds.

So part of the equation in this new environment is the difference of voice. I wrote a piece for the Industry Standard called Fear and Loathing on the Web. Reacting to that article David Weinberger said: "the dogs have it right. They want to take a good long wiff. Companies that cannot, or will not, speak in a human voice built Web sites that smell like death." That was the beginning of Cluetrain and its quoted right on our home page next to the run-over armadilo that is road kill. We recognize each other by the kind of voice that we used to communicate human being to human being. We do not recognize corporate rhetoric as belonging to our conversations.

And there's another model here as well. Netscape started out with that kind of voice. But then they hired Barksdale who came in with the pin stripe suit spouting the corporate ethic and went out and tried to sell servers. And in doing so Barksdale just turned off this religiously committed market that wanted to have a conversation with a company that was speaking its own language. He turned it off like one would a light switch. Now Netscape wants to blame Microsoft for undermining its business. They killed themselves because they didn't have the nerve to walk the talk.

Tearing Down the Berlin Wall Separating Work Force from Customers

Now let's go to the other side of the metaphorical firewall that separates companies from the marketplace. On the other side, inside the company, there are intranets with the same TCP/IP technology that belongs to the Internet. How can you tell the good ones? They are likely to be good if the company does not have a fascistic top-down intranet application with HR manuals and all that kind of junk. And if the people who are using it are also the people who are building them. [SEE ACCOMPANYIONG TEXT BOX ON INTRANETS.]

We say in the manifesto, if you look at the conversation going on in an open healthy Intranet, it bears an uncanny resemblance to the conversations of the marketplace. One of the Manifesto's statements says: if you want to sell to me, get down off your camel and take your shoes off at my door. Now this may be a little bit jarring, but there is a legacy reason that it is there. At one point I had started to write a narrative version that would be in a column to the right of the 95 theses. It started with a hypothetical market in Mesopotamia 5000 years ago. Then the real marketplace was almost certainly in the midst of the town square. It was wherever people who lived there went there to talk.

COOK Report: Yes, and what happened that attracted people like you and I and Dave Hughes to this technology some 20 years ago, was a sense that just maybe a return to something like this marketplace was an inherent possibility in the maturity of this technology.

Locke: True. We call this part: ancient markets. People did not go first and foremost to these old markets to buy olives and things like that. They went there to hear stories. And because these were guys coming in on camel caravans from God knows where there were a lot of stories to be told.

COOK Report: Twenty years ago on the source, one of the earliest commercial networks, Dave Hughes discovered that storytelling held to the key to getting other

Two Styles of Intranet Building

There are two styles of doing Intranet's. One is the bottom up approach where you get people who are really focused on skunk works projects, and on getting mind-share and buy-in to be used in building consensus for things that they are really turned on about. Now this style, driven by true enthusiasm, is really like the human voice that we talk about in our manifesto. But you also have the other kind of style. This is very much top-down, very much like Lotus Notes only running on TCP/IP. It is run by the "happy-talk" of the corporate PR department. Its idea of creativity is to make available to everyone things like the Human Resources manuals and the cafeteria schedules. It is one where the companies process their employees. And it is one that does not fly very well because people feel, and rightly so, that they are being subjected to broadcast. Attempts to do this within the organization have been fraught with command and control issues: do this; don't do that and so on.

Now if you want to talk intranets, I've got a good war story there. Back in '94, when I was doing MecklerWeb, Dun & Bradstreet was one of my core partners. I would be invited then on regular e-commerce sessions that they would have with IBM and other large clients. There was a fellow there named Ted Wolf who was head of the IT group and who managed Dun & Bradstreet Information Systems. Ted gave a presentation to this group right on the heels of a fellow from Lotus Notes who had given a really slick PowerPoint presentation. Ted stood up and said: let me see if I have this straight. this will cost 500 bucks a seat, right? The Lotus Notes guy answers: at a minimum. Ted continued: and it's going to need corporate buy off so an entire IT study will have to be done on needs and requirements.

And then Ted looks at his watch and says: let's see. It's 3 clock now. I have a pretty small unit with D and B — some 350 people. You know, before the end of the day here I could download the CERN Web server. Have it in place. Put HTML templates out to everyone. Send some email. And he goes: you know it wouldn't have all the bells and whistles of Lotus Notes, but when people come to work on Monday morning, it's in place and its free. I think you have a real selling problem with Lotus Notes. (IBM had recently bought to Lotus Notes and I was just beginning to realize they'd bought it at the end of it's lifecycle. Of course, being IBM, it probably took them another two or three years to realize this.)

But Ted wasn't finished. "This is in fact what we've done," he concludes. He then goes on to give a presentation about how a core team in his group got really excited about the Web. They knew nothing about it. But they went from zero to 180 and built an intranet that was sucking information from all over Dun & Bradstreet; filtering it; putting it together. With no previous experience or background they did the whole thingin only six weeks.

Unfortunately right around then the notion of intranets caught the attention of the corporate higher ups. Suddenly it was: "here come the suits." At first Wolfe's people thought that the suits would give them the resources they needed to take it the rest of the way. After all, they were gung-ho — working 18 hours today because they owned the project. But instead of supporting them the suits brought in the lawyers and marketing people. Suddenly, he said, it was no longer "ours". They took away from us and we couldn't do anything without six levels of approval. Everyone became totally discouraged and bailed out of the project because it was no fun any more. And in this meeting, right in front of me he is talking to them face-to-face and saying: you know you wrecked the best thing that we'd ever done. You wrecked it because you came in and took it away from us and put it in the hands of people who don't understand how it works net users to corporate with him.

Memes and the Marketplace

Locke: In the Internet possession of the story is preeminent. If you really look with fresh eyes at what happens on the Internet, you will see that people use a very large portion of their time in trading jokes. This is analogous to the stories of the ancient marketplace. Such stories often express what we call memes or common ways of thinking about problems and events. Last fall the meme that probably travelled the fastest in the history of the world was ‘when John Glenn comes back from space lets all wear ape suits.” As soon as he went up in the shuttle last November some wag said it and the first time I saw it, I thought that I was going to swallow my tongue. I had just sent it out to 3,000 people that morning on the my EGR list saying: this is really good, pass it on. That evening, coming back from Denver, I had the radio on and it’s All Things Considered talking about the Glenn and how he is doing up there. And then adds: there is an interesting thing going on in the Internet. And one of the Cluetrain guys David Weinberger unbeknownst to me comes on and says: “however on the Internet it has jokingly been decided that when John Glenn comes back from space we should all wear ape suits.” No corporation could get word around with the speed at which the ape suit meme traveled — not even by buying a Super Bowl ad.

Another meme was Hank, the angry drunken dwarf. People Magazine has a contest to select their annual Person of the Year. Now last year they decided they would do it on the Internet. The issue was who is your favorite person? Take from our lineup of movie stars and other celebrities. Well some person on the Internet designed it to nominate a fellow who is the sidekick of Howard Sterne. This sidekick has the name of Hank, the angry drunken dwarf. I had never heard of him. But he started getting a block of write in votes. The idea that Hank should be supported and instruction in the means of doing so started traveling the net. Word quickly became: write in Hank the angry drunken dwarf. In short order Hank was at the top of the list. People Magazine decided that this was not what it had intended and announced they would remove the Hank votes. Unfortunately for People, by the next morning, due to the efforts of his behind the scene supporters, Hank was on top again. People would delete the votes refusing to acknowledge them, but within hours they would be replaced. This was an example of a community of people just playing with its power.

They are saying where are the edges? Where are the buttons? How does this work? When this gets organized, that hockey stick I was talking about is going to take off. It goes right through the roof. It can bring down Microsoft; it can bring down General Motors if it doesn’t like their friggin bow tie. That is power.

My experience with MecklerWeb was five years ago. A lot has happened since then. The same dynamics are still there. Only today they are orders of magnitude more powerful. But no one who really understands these dynamics is talking about them — except for little e-zines out on the fringes of the net. JoHo and EGR are e-zines written by people who live and breathe this stuff but who are not being invited to the big mainstream conferences to talk about it. We just decided that the time is ripe for this because all these forces are coalescing. People are talking about trillions of dollars in e-commerce, but we think that they are just simply deluded because they think the metrics for measuring e-commerce are things like click through rates and how many things you buy given an amount of time online. Those who would do e-commerce assume that there is no change to be made from what they would do in advertising in the local paper or putting ads on television. They take the same heuristic’s, the same algorithms and extend them into the net. And suddenly you have the latest figures from Foerester: by the year 2002 there will be “x” trillion dollars in e-commerce.

Our messages are easily brushed off in the same way that Borders and Barnes & Noble brushed off Amazon. But ignore this one at major potential peril. This market is getting really smart. It is playing with it and having fun. It’s very likely to reach point where a company may send out a press release which is so bad that the market will turn on the company and take it down. Remember the movie “Network” when the guys says: “I’m mad as hell and I’m not going to take it anymore.” That’s our message. We would love to come away with millions of signatures on Cluetrain which we could then show to the corporate world and say: this is frontline market research that you’re not getting from Gartner, or Forester, or Giga, or Datapro or any of those guys. We want to put it in their face and say: understand this or die. Actually we think the Manifesto is rather profound and that it will gate how up to two or three trillion dollars in e-commerce flow over the next three to five years.

If We Are Right, Empires Are Going to Crash

Our question is: yes, but who will be the beneficiaries? Here is a binary way to view Cluetrain. If we are wrong about these things, you can laugh at us. We will have had no impact and will have been just a little blip on the Internet — a bunch of guys jumping up and down and waiting their hands around. And who cares? Because we were wrong. But, if we are right, empires are going to crash based on what we are saying, and new ones are going to rise. However it is unlikely that the Amazon model will be repeated 1000 times. I think much smaller companies will be replicated — millions of times. You know three or four or five or six people who can make a really good livelihood with a very small niche.

COOK Report: And, as long as you have a ubiquitous and resilient communications system like the Internet capable of connecting all those niche operations, the economy not only survives but prospers.

Locke: The economy not only survives, it is a lot healthier. Today you may have a lot of companies converging through large mergers but the process of their doing this speaks against the human voice that powers Cluetrain.

A lot of people who read the Cluetrain manifesto will scan through the individual statements and find that they agree with most all of them. But if you also try to read it a bit more slowly and really critically, I think you will find that the pieces add up to a lot more than 95 simple statements.

COOK Report: What they do add up to is what we’re been talking about right here. Remember Soshanna Zuboff’s mid-’80s book called In the Age of the Smart Machine? They are she became the first person to point out how computer systems could threaten and do away with middle management. When we are seeing now than at the end of the ’90s with the success of the Internet is the Internet doing away with the intermediary between the creator and the consumer.

Locke: That’s true and people refer to it as disintermediation. Trying to predict cause and effect for all this is difficult but likely to be worthwhile. Consider the meme that underlies those of Cluetrain. It is the question of the relationship of the Internet global economy. I am saying that the Internet did not drive the forces making up the global economy because those forces were already in place when the Internet came along. The Internet basically served as a catalyst that speeded them up and glued them altogether.

The message of Cluetrain is also gluing together a lot of otherwise disparate views. They range from those of Eric Raymond, the anarchist-oriented founder of the open software movement to those of Ruth Perkins CAIS, Florida Department of law-enforcement who wrote in her Cluetrain sign up: “thank you for solidifying all the thoughts and mission I’ve had for so long. I am a wholehearted signer and practitioner of your manifesto.” The Cluetrain Manifesto is a message with which the whole family can play.

Executive Summary

Juniper Networks, pp. 1-8

Tony Li, one of the designers of the new Juniper Networks terrabit routers, in his first interview, explains the major issues and innovations in the design process and discusses the problems of intelligent traffic management at the core of a "stupid" Internet.

After its initial stage of VC funding, Juniper partnered both with several carrier and equipment companies: Ericsson, Lucent, Nortel, 3Com, and Siemens/Newbridge, and with end-user investors, such as AT&T Ventures, the Anschutz Family, and UUnet. The partnering gave them early feedback during both the design and implementation phases of development. The partner's investment in Juniper also helped to cement the partnership, ensuring a mutual commitment to the process of bringing the M40 to market.

Juniper has built a router that separates routing computation from packet forwarding by carrying on the two functions in parallel. It has also taken advantage of custom made ASIC (Application Specific Integrated Circuit Design) to build a router that achieves a look up rate of 40 million packets per second with a single chip (ASIC). Juniper's innovation in its switching fabric is in an increased efficiency and ability to use less sophisticated technology to implement the same size fabric as compared to other, traditional crossbar designs. It gained an efficiency that translates into fewer parts and a more efficient utilization of the parts that are used. The result is less costly less to the end-users because it's less costly to produce. The use of a single system wide buffer also holds down costs.

The M40 makes significant use of MPLS which provides a hybrid network architecture, where one can support, at the same time, both datagram mode forwarding and a connection-oriented service. It offers the hybrid strength of being able to route packets through the network based on something other than the destination address. This gives the network an major degree of increased flexibility, which makes it possible to provide new services such as voice over IP and VPN services on an integrated Internet backbone.

Unfortunately, there's a conflict between the amount of separately defined QoS levels one wants for customers at the edge of the network and what one can actually aggregate and deal with by routing on the backbone. This is being dealt with in two ways. First is the application of MPLS to aggregate flows with QoS and routing properties. The second approach is the work that is being done in the Diff-Serv working group in the IETF. Diff-Serv provides a convenient and easy way of "coloring" packets with particular levels of QoS requirements. These are bulk, generic kinds of "coloring" that are not flow specific and thus require minimal amounts of state within the router forwarding function.

The current Diff-Serv proposal supports up to 64 different "colors", or more precisely, 'Per-Hop Behaviors' (PHBs). Some of these are defined to be global, some are reserved for local use. It's left to each provider to determine the PHBs that are applicable for the services that they wish to deliver. However, because the definition of a PHB can in fact be wholly local to a provider, it raises an interesting question about the ability to define and deploy interprovider Diff-Serv functionality.

Differentiated services within a single provider's network is of interest to the provider, but to the end user, who wants to use the Internet backbone as a global facility, differentiation with a limited scope isn't a practical solution. Some providers are expecting to differentiate themselves based on their domain-specific services and are not interested in supporting globally defined services. This would present a drag on the deployment of global services that could be overcome only when there is a de facto standard and the non participants are the exception.

It is possible that these scenarios can be avoided, but it will take initial experimentation with interprovider QoS agreements, in both bilateral and multilateral agreements. An appropriate forum where such issues could be aired and global service definitions discussed without antitrust problems is needed.

vBNS Tests Juniper, pp. 9 - 15

In 1995 the NSF funded MCI to provide a very high speed backbone service to connect the national supercomputer centers. Given the administration's interest in promoting Internet2 and the NGI, the vBNS effort has now moved to a focus on prototyping the next generation internet backbone. We interview Rick Wilder Director of Engineering for the vBNS.

On part of the vBNS Wilder is testing RSVP sessions mapped onto ATM swithed virtual circuits. Wilder is also using RSVP in setting up label-switched paths through the backbone. The application here is traffic engineering, For as traffic conditions in the network change, it's possible to easily change the loading of paths on the fly.

Wilder expects to see decreasing use of ATM on major internet backbone not just because of the well known cell tax but also because the amount of complexity in the router interfaces needed to do the segmentation and re-assembly, i.e. the conversion of packets into cells and back again. The complexity is actually in the hardware and/or firmware and this, in turn, requires that the interfaces themselves be more complex and more expensive. When MCI went to OC3 routers based on ATM, it took longer to get reliable router interfaces than it did with DS3 interfaces or with Packet over SONET (PoS) interfaces which is what it is currently running.. With PoS they will use MPLS, at least initially, to replace the traffic engineering capability they currently have with ATM..

Having begun tests on the Juniper M40 in their labs last summer, they started to use them to run production traffic on the vBNS in early February. They have not noticed any problems arrise from the trial and are generally happ with the way things have gone. Wilder points out that DWDM could be used to provide QoS. He adds that ISIS (Intermediate System to Intermediate System) routing protocol is probably the protocol where the most QoS routing work is now being done in the IP world at the IETF. That's because its being used by most of the very large ISPs for use in their backbones today. The IETF working group chaired by Tony Lee is doing a lot of work in the area of adding QoS differentiation to the routing.

He finds that QoS in a connectionless Internet packet network can be achieved. It's just that conceptualizing the problem so far has been a daunting task. And for very good reasons. Breaking it down into its component pieces and getting enough different efforts working together to build the whole architectural solution for a problem as complex as this one doesn't come quick and easy.

E-Commerce, pp. 16 - 22

We review a study on e-commerce "Portals To Profit: E-commerce Business Models and Enabling Technologies, published on April 16 and available from www.datacommresearch.com.

According to one of the co-authors: "Our main thesis is that e-commerce competition will transform all commerce, destroying many if not most traditional business models, and forcing companies to invent new ways to make money. Thus, it is imperative that businesses understand the new models, many of which involve cost-based pricing, below-cost pricing, auctions, reverse auctions, ad-targeting, etc."

So far so good. But the report goes on to paint its picture through the eyes of the Fortune 500. E-commerce is primarily about size and, as such, is a race to see how the large corporations can translate their standard views of the world into this new medium. The views are basically the old industrial age mechanics of economies of scale using, this time, the latest digital technologies and suites of software agents behind the scenes to manipulate and shape the thoughts and response of the customers. The customers are portrayed mechanistically as sets of eyeballs on which the new technologies act.

It certainly represents a legitimate point of view and its focus on automated mechanistic ways of dealing with the masses describe a modus operandi that may be all that is needed for success. After all 90% of the people on the net have been there less than 2 years. They may never be able to see the Internet as much else besides TV with a buy button. However another view exists. The Cluetrain manifesto suggests that the Internet makes a whole series of special relationships between customer and businesses possible. It is the view of Chris Locke. Someone who has, as have we, been on the net for nearly 20 years. We believe that, the longer one is on the net, the more the experience broadens one's horizons.

In this world view the web is driven by the corporation's customers who can join together to route around companies that can't or won't

Continued on page 24

Continued from page 15

can be solved just by throwing more bandwidth at it. Enron is doing this with its overlay network where they just bring lots of additional bandwidth to Real Networks for Real Audio and Video. On the other hand, since new uses for bandwidth appear to be generated endlessly, it's likely that Enron will find its pipes filled up sooner than it thinks and will need QoS applications itself at some point.

Wilder: It's a good question but I'm not familiar enough with Enron's infrastructure to give a detailed answer. But I do know that you need some kind of control in the signaling from the application so that the application can tell the network what kind of service it needs. The reason is that there will be higher speed access into the network with things like DSL and bandwidth will become cheaper. But we'll never have a homogenous environment. Suppose you have a high bandwidth web server, probably DS-3 connected, trying to feed a dialup modem user. That link can easily be overloaded and cause all kinds of problems. A few years down the road it may be a DSL user but he may be talking to an OC12-connected streaming media server or something. He's potentially going to have exactly the same problem we see today with the dialup modem user. It will just be at a higher bandwidth—but it's the same bandwidth mismatch. So you need controls in signaling to gear the network and the two ends of the connection to provide the right service.

Keep in mind that applications are getting more and more complex and using different amounts of bandwidth. To get all of these factors properly cooperating with one another isn't easy. If someone is provisioning the backbone, for instance, they can over provision it and not have a problem there. But that moves the congestion out to the edge of the network, to the access links or to the host themselves, you'll find these devices running out of compute cycles and they won't be able to give you assured end-to-end performance.

Continued from page 23

meet their needs. Web enabled consumers, according to Cluetrain, are there not to be manipulated but to join together in efforts that will replace those companies that: "don't get it." Cluetrain represents both a culture and ways of looking at Internet commerce that business would be very much ill-advised to ignore. Locke had the following to say in a lengthy interview with the *COOK Report*. My experience with MecklerWeb was five years ago. A lot has happened since then. The same dynamics are still there. Only today they are orders of magnitude more powerful. But no one who really understands these dynamics is talking about them — except for little e-zines out on the fringes of the net.

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

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