new world catalysts

high-level shifts in the networking market

BY JUDITH ESTRIN

The Internet—or, more broadly, the use of Internet Protocol (IP) technologies in everything we do—is changing the way we work, live, play, and learn. Over the last decade, the Internet has had its most profound impact on the first area: the way we work. Even in the higher education market, the Internet has been used primarily in how one works to deliver education as opposed to delivering the education itself. Yet as we think about the next decade, we are looking not only at continuing to change the way we work but also at changing the way we live, play, and learn. Some of the most exciting things happening in the networking market are aimed at recognizing that potential. But to do that, we have to change our whole mindset.

Networking used to be very tactical, concerned with wires and protocols—for example, token-ring versus Ethernet or IP versus ATM. We have moved from thinking about bits and bytes and wires and protocols to thinking about how to connect users to services. Today networking is about providing information, and it's about ubiquity. Because in the end, the technology is just the means to the end. It is the content that matters. What we need to do is get the Internet to the place where it again becomes transparent, so that we can go back to focusing on the content.

I would like to discuss four high-level catalysts driving this change in our industry. The first is convergence. We seem to talk about nothing else in the industry today—the convergence of data, voice, and video into a single infrastructure. The second catalyst is the increasing intelligence of networks to provide the services required. The third is mobility, which changes the dynamic and the use of the network, and the fourth is home and consumer networks.

Judith Estrin is Chief Technology Officer and Senior Vice President at Cisco Systems. Before joining Cisco, she was President and CEO of Precept Software, a multimedia networking and software company that she co-founded, was President and CEO of Network Computing Devices, and was a co-founder of Bridge Communications, a pioneer in intranetwork routers, bridges, and communication servers. Judy also serves on the Board of Directors of FedEx Corp., Sun Microsystems, and the Walt Disney Company.

Convergence

I think it is fair to say that convergence is basically accepted now. This wasn't the case a year ago, however. A year ago it was somewhat controversial as to whether we were going to see a single infrastructure that ran data, voice, and video together. We now know that we will see this; it's just a question of getting there.

The new infrastructure is being driven in two phases. Most of what we see today is focused on voice and data, the first phase. This is because the first phase of convergence involves the merging of the telecommunications industries and the data communications industries as service providers look to upgrade their infrastructure. However, once that infrastructure is in place, and once we have ubiquitous bandwidth, we will enter a second phase, which is all about content. The focus then will broaden to include video, because we're talking about the convergence of the entertainment industries and the information industries. This will probably happen within two years as we start to see, again, enough critical mass of bandwidth available to the home.

One of the things that we need to realize as we move to a converged infrastructure is that it means doing some things in different ways. The telephony world, the circuit-based world, has historically done

Illustration by Glenn Pierce

things one way, with certain business processes. It is typically a vertically integrated proprietary world. The IP world has been based on open systems and standards and interoperability between many vendors. One of the more interesting questions is how we can converge the cultures and the business processes, and not just the technology, of these worlds. Actually, the technology is at times the easier part of this convergence.

So how do we integrate data, voice, and video to implement applications in new ways? It is not just about replacing a telephony infrastructure. In the enterprise, the decision about whether to go to a converged infrastructure and replace a PBX involves more than determining whether voice-over IP will work as well as the PBX does today. The decision should be driven by the advantages of applications that integrate data, voice, and video, such as new customer-care applications.

We draw connections into the cloud, and we have clouds that are independent of each other. The whole Internet culture, the Internet architecture, was based on independent networks that all hooked together. This distributed architecture is such that failures in pieces of the network will not bring down the entire network. So we have distributed intelligence. This cloud-based approach is what I call "architecturally agnostic." What I mean is that one of those clouds can actually have strings in it. A subset of the network can be ATM based, for instance, or point-to-point based. But the intrinsic architecture of having packet-distributed IP end to end gives a level of reliability and scalability.

One of the important aspects here is that growth today is different from what it was ten or twenty years ago. We used to be able to predict growth. Today, because we are using the network and computers in so many new ways, it's almost impossible to predict growth. New

With voice services—actual call set-up and tear-down, billing and accounting, security, or quality of service—we must be sure that we don't just take the services from the telephony world and pluck them down on the IP world.

What about reliability? Can an IP-based infrastructure provide the reliability that we need and that we're used to in the telephony world? In fact, IP networks are intrinsically more reliable than telephony networks. Let me explain. Think about the old world, the telephony world, as being about strings, about point-to-point connections. When we draw telephony networks, we draw point-to-point connections, "strings," with very large centralized switches and processing that make all the decisions for hundreds and thousands and in some cases millions of transactions. A major focus in the old world has been on hardening the centralized systems, because there are many single points of failure.

On the other hand, in the IP world, the data world, we draw clouds. We don't draw point-to-point connections. ".com" companies are having to add capacity at four times their predicted growth because of what they see in terms of spikes and new users. So the growth rates in this industry, whether for a ".com" company or for an IP shop internally, are not predictable anymore and are growing at a much faster rate. As a result, we need a much more scalable architecture.

Then why is it that picking up the telephone today is still more predictable than using an Internet connection? This reliability, or predictability, is not intrinsic to the telephony architecture. It comes from years of maturity of the telephony system. We still have a set of opportunities or challenges being addressed in the IP industry. Some of them are business practice, but some are technological. For instance, if a router goes down, another path is found very quickly. That's called *routing convergence time*. The convergence time needed for file transfer is very different from the convergence time required for voice. If a router goes down when you're in the middle of a phone call, the convergence time needs to be much quicker than for a file transfer.

One of the most challenging issues we face is rebuilding services around a cloud-based architecture. We know how to do things in a point-to-point architecture. But with voice services—actual call set-up and tear-down, billing and accounting, security, or quality of service—we must be sure that we don't just take the services from the telephony world and pluck them down on the IP world. If we do that, we introduce strings on top of our clouds. And we introduce single points of failure, which is exactly what we don't want.

Finally, I would like to point out that there is still a critical role for ATM in a cloud-based world. The important role that ATM plays is not as an end-to-end architecture but as a connection point between a string-based architecture or a point-to-point, circuit-based world and a distributed IP world.

When is this all going to happen—and at what speed? The service providers see a migration path. Today many of the service providers are using IP technology to off-load the data traffic from their circuit networks. They are running converged networks, or so it looks to the outside world, but their internal networks are two side-by-side networks. The next step is to start introducing IP telephony into the systems so that new applications that require voice-over IP run through a converged network, but again still side-by-side with the public network, which is running the dominant portion of the voice traffic. And then finally, most voice will move over to the converged network.

What is the timing of this migration? The middle phase is just beginning: people are working in the labs and starting to offer voice-over IP services for specific applications, like long-distance across continents. In my opinion, this middle phase will probably last for several years. But the important thing to remember is that this is a migration, not a one-step process. It is an evolution of steps. And it is inevitable.

What about educational, government, and commercial enterprises? Many enterprises today have already moved to a common wide area network. The second step is installing switches that can run IP telephony in addition to data, and the final step is where you see the real benefits, as you migrate to applications that take advantage of converged data. One of my favorite such applications is customer care—whether this involves registering students for

Operational automation is critical to scaling the network; otherwise, managing the complexity of all these different users and services is just too hard.

courses or helping customers with technical support. The idea is for the customer to use Web-based self-help and not have to talk to a human being. If I am a customer today, I go to the Web, and if I don't get what I want, I pick up the phone. But when I pick up the phone, I start from scratch because the person on the other end of the line is not integrated in what I've been doing over the Web. On the other hand, if customers go to the Web, get stuck, and can push a button and converse with somebody who knows where they are in their Web-based session and can lead them through the necessary Web pages, you better believe that next time they'll go back to the Web. They won't go directly to that human being, and that's the behavior you want: to get people off the phones, resulting in better productivity.

Increasing Intelligence of Networks

Connecting computers together is much easier than connecting users to services. Connecting people requires another level of capability in the network. The required level of things like security and personalization changes. So we need to think about networks as more intelligent. An interesting architectural battle involves whether to build that intelligence into the network or whether to build that intelligence into the client's servers and keep the network relatively less sophisticated, or "dumb," as some people would say. I believe that the intelligence should be distributed in the network. Note this does not mean that each device in the network becomes more complex, but that the network consists of simple switching or routing devices coupled with more complex network servers providing added functionality. The last decade, whether everybody liked it or not, was about Windows and PCs. That is changing. I'm not predicting the demise of the PC, but more and more we're going to see an increase in network appliances or dedicated devices. PCs will remain important for general-purpose processing, but dedicated Web browsers and pagers or cell phones that become Web phones will change the equation. The device dedicated to one or a couple of functions can be smaller, more reliable, and cheaper because it is dedicated. And as functions are

stripped off the device, some of that intelligence can be put onto the network servers. So the network needs to become more intelligent to serve the devices as they become more specific.

What do I mean by *intelligence*? This means making the network more user aware and more content and application aware, differentiating between different types of applications or content, and having different behaviors depending on the user. Operational automation is critical to scaling the network; otherwise, managing the complexity of all these different users and services is just too hard. So whereas network management years ago was a nicety, and started to become a necessity as networking became more critical to what we do, it becomes even more of a necessity as we look at building more intelligence into the network.

The types of things I'm talking about are quality of service and voice, video, and security services. These are all services that need to be provided by the network. An analogy may help to explain. The idea of the network providing more intelligent services is very similar to what Federal Express did twenty-five years ago in the transportation industry. Before Federal Express, you could give a package to UPS or the post office, and it would get to its destination. The shipper had to worry about things like confirming receipt. Similarly, in the networking world, you put data in an IP packet, and it pops out somewhere else. But FedEx added a set of services on top of just the transportation of goods. Whether it's billing and accounting or marking packages for special handling or time-definite delivery, all of those services are performed within the network. You hand the package to FedEx, and FedEx takes care of all of it before the package gets to its destination. FedEx does this by having a very sophisticated network of couriers, planes, and trucks overlaid by information systems. If you take the information systems of FedEx and put them on Airborne's network, you won't get the same capability. If you take Airborne's IS systems and put them on FedEx's network, you won't get the same capability. It is the coupling of the software and the delivery network infrastructure that provides this intelligence. It's the same in the networking world. To provide the necessary services, you need capabilities in the boxes themselves, the switches and routers; you need special-purpose devices—cache engines, fire walls, video servers—what I call "mission-specific network devices." Then you need the policy and control services, all of this tied together by directory services.

Mobility

The third catalyst is about mobility. Clearly, the mobile user is no longer a second-class citizen. Whereas we used to talk about the special needs of traveling salespeople, today everybody in an enterprise is mobile because of telecommuting, because of traveling, and because of cell phones. Thus we need to stop thinking about mobility as an afterthought. In the networking world for the last twenty years, there's no question that mobility has been an afterthought. We now need to think about the mobile users as an integral piece in building our networks.

Often people say "mobile wireless" and forget that mobility is also about the services. People need similar services as they move from place to place.

Part of this is about wireless. We're seeing new technologies and new techniques in many areas of wireless. As I look at start-ups today, some of the more interesting developments involve new wireless technologies. In addition, large companies like Motorola are embarking on third-generation cellular mobile systems to bring more bandwidth.

But it's not just about wireless. Often people say "mobile wireless" and forget that mobility is also about the services. People need similar services as they

move from place to place. We should think about mobility for connected users and not just wireless users. The state-of-the-art today is focused on mobile IP, which provides a persistent IP address, but I believe that over time the focus will switch to persistent services. There's no reason why a person's IP address can't change.

We also need to realize that for a long time to come, mobile bandwidth is going to have lower capacity than fixed bandwidth because it's more expensive. Thus special user-interface requirements will be needed and should be included in the design from the beginning. Everybody is so excited about these fancy Web sites that have huge amounts of graphics, but these sites simply can't be accessed from some of the mobile devices. In addition the mobile devices often have small screens, so the look-and-feel of the user interface needs to be thought of differently.

The new trend will be not just network appliances but networked appliances. The network will connect the refrigerator, the washing machine, the microwave . . .

Home and Consumer Networks

The fourth and last shift involves networking to and within the home. We have talked about home networking for a long time. Over the last several years there has been an incredible change in terms of the cost of networking technology and of the requirements from home users, both of which will make home networking a reality over the next few years.

We're now seeing portals, whose main job is to make the Internet usable for those outside our community. The portal is bringing the Internet to the masses and making it easy to use. The whole idea is to make our lives easier. Today, however, the Internet still complicates a lot of people's lives. They spend hours and hours and hours trying to figure out where information is. Over time, portals will change this.

I already mentioned network appliances. There are all sorts of new electronic devices that need to be connected to each other as well as to the network. In fact the new trend will be not just network appliances but networked appliances. The network will connect the refrigerator, the washing machine, the microwave—not so that the microwave will be able to "talk" with others but so that the repair people will know what's wrong with a broken microwave and may even be able to fix it from a remote location.

Another important trend is communication to the home. People ask: what's going to win—DSL (Dedicated Service Line) cable or wireless communication? My answer is that they're both going to win or neither is going to win. We can't look at this as an either/or. The important thing is to get the bandwidth out to the home and to design your services so that they are not fixed to one or the other because depending on where you live, one is going to be more effective or cheaper than the other. The biggest

mistake you can make is tying your content to a single mode of distribution.

What are the challenges for consumer networking? Ease of use is a critical one. Networking technology is not where it needs to be in terms of being easy to use and easy to manage. And we need more scalable deployment.

On the other hand, being "always connected" changes the pattern of your life. When you have an always-on connection, you can use the Internet in new ways. The chief technology officer of one of the cable companies once told me: "Look, you don't make an appointment with your refrigerator to get a beer. You just go into the kitchen, and you pull out the beer. Well, when you have an always-on connection, you say I need to order that book, and you go get it." You don't have to make a note to remember that the next time you dial up to your connection, you need to do a list of things. The Internet is used much more often with an always-connected system. However, being always connected also puts new requirements on security services. We must provide products that have the ease of use required for home markets while still protecting consumer privacy.

As we look to the future, networking will go beyond helping businesses and enterprises enhance their work operations, to helping us change the way we live, play, and learn. Networking is about new types of content, whether e-learning or video content. It is about new types of commerce: the Amazons and the E*Trades. And it is about community, about connecting people, independent of geographics, who have not been hooked together before, whether these are people with a certain illness, senior citizens, or kids who go to their computer and interact with their friends and play their games online. The networking paradigm shift that we've seen in businesses and enterprises is now moving out, affecting all our lives in numerous new ways.

This article is based on a speech delivered at the Seminars on Academic Computing conference, Snowmass, Colorado, August 10, 1999.