

# Demands Shift For Next-Gen Switch/Routers

Bart Stuck and Michael Weingarten

## Enterprise networking trends will drive demand for more flexible and powerful switch/routers.

In the late 1990s/early 2000s, speakers at the Next Generation Networks (NGN) conference used to talk about a next-generation network organized around the following principles:

- It would be largely IP at the core, with any remaining TDM traffic encapsulated in IP.
- To facilitate differentiated quality of service (QoS), it would use MPLS/DiffServ.
- To facilitate video, it would use IP Multicast.
- For security, it would use IPSec.

Apart from these specifications, it would be a relatively dumb network connecting intelligent edge equipment. To that end, network equipment providers were expected to develop 10-Gbps chips and boards that would support traffic growth rates of 10X per year. There also was considerable interest in multiprotocol chips and boxes that would accept TDM and IP data flows and would uplink via IP.

It's now three years and one depression later. What happened? At the network core, some people actually came out with working 10-Gbps silicon, but they achieved limited market penetration. At the edge, we're just starting to see general-purpose switch/routers handling 10/100/1000-Mbps copper ports for enterprise LAN applications, with a few 10-Gigabit Ethernet ports for uplinks. These boxes support Layer 3 and 4 traffic management, while higher-layer processing has been confined mostly to the front-end traffic handling in Web server farms.

From this description of things, the general sense you might get is: OK, we've lost a few years. Just take those circa-2001 deployment forecasts, re-label them 2005, and deploy the networks people were predicting before the crash. We were correct about what would be needed; our timing simply was off.

As the authors look at the present situation, however, we don't agree. Sure, we're going to need higher-speed IP networking equipment with improved QoS, security and multicasting capabilities. However, conversations with more than 100 colleagues, active in the enterprise networking space, have made it increasingly apparent that the next-generation network for 2005 and beyond will need to address a wider range of networking requirements than were on the radar screen in 2001, as shown in Table 1. We have identified seven of these, specifically:

### Executive Summary

Next-generation switch/router designs developed before the so-called telecom winter won't meet tomorrow's enterprise networking needs. Instead, switch/routers will have to address a wider range of networking requirements, including:

- Changing traffic patterns
- Layer 4-7 Application Awareness
- Security
- Wireless
- Storage
- Grid Computing
- XML

Developments within each of these technology areas did not stop with the slowdown in tech and network spending. New switches/routers are needed that can:

- Handle these technology developments in the enterprise LAN rather than in the WAN.
- Scale to treat them across enterprise core and remote sites.
- Read and differentially process heterogeneous packets at wire speed
- Read and convert among multiple protocols at wire speed.
- Accept remote programming changes for tighter security.
- Scale to very large non-blocking capacity levels.
- Handle very large packets.

Do such switch/routers exist? Not yet, but they need to □

*Bart Stuck (barts@signallake.com) and Michael Weingarten (mikew@signallake.com) are Managing Directors of Signal Lake, an early-stage telecom venture capital fund (Westport, CT and Boston, MA)*

**TABLE 1 Switch/Router Requirements For Different Applications (High = XXX; Low = X)**

	Requirements	Located at the LAN	Read and differentially process heterogeneous packets on the fly	Multi-protocol at wire speed	Can accept new programming remotely	Scalable to very large non-blocking capacities	Can handle very large packet sizes
Applications	Changing Traffic Patterns	XXX	XXX	XXX	XX	X	X
	L 4-7 Application Awareness	XX	XXX		XXX	XXX	X
	Security	X	XX	XX	XX	XX	X
	Wireless	X	X	X	XX		
	Storage	X	XX	XX	XXX	X	X
	Grid Computing	X	X	XX	XX	X	X
	XML	X	X		X	XXX	XXX

**Changing Traffic Patterns**

Since 1999, the mix of on-site and off-site data traffic has begun a fundamental shift, with an ever-increasing volume of traffic leaving the LAN and going onto the Internet. Examples abound, ranging from business Web portals to consumer shopping and file sharing. This shift will have profound implications, as profound as the shift in the mid-1990s away from multiple, heterogeneous, departmental LANs to the hierarchical Ethernet/IP enterprise network architecture. Unlike that transition, which changed premises data networks, this shift will change the requirements for switch/routers at the enterprise edge.

Currently-available Layer 2 edge switches can handle tens of millions of packets per second, with client access at 1 Gbps to 10 Gbps, while Layer 3 routers can handle hundreds of thousands of packets per second. The more traffic needs to leave the enterprise, however, the less likely Layer 3 routers will be able to keep up with the offered load, especially while delivering wireline QOS with minimal input buffer overflow.

The prognosis is actually worse than this, because Layer 2 edge switches also will need Layer 3 and 4 awareness to handle QOS, and this extra processing will affect their ability to handle 1 to 10-Gbps offered loads with minimal latency. If Layer 4-7 processing is required, the problem gets even worse!

The solution is next-generation switch/routers designed to handle higher percentages of off-LAN traffic, with full Layer 3-7 awareness, at higher speeds. We're going to need new chip architectures and new network equipment designs, much as happened in the mid-1990s when the hierarchical switch-then-route enterprise networking design took hold.

As was the case in those days, switching at wirespeed rates will be the mantra in times to come. This time, though, wirespeed will have to apply not just to switching, but also to full packet processing for higher-layer awareness.

**VOIP And Video On The Switch/Router**

The 1999 approach to mixed-application networks called for intelligent edge devices (PCs) to code packet headers with application-based priority fields. Then L2 switches at the edge and L3 routers in the core would read these packets on the fly and give priority to applications like voice.

Arguably, in the past few years, what we've been seeing is more of an "overlay" approach, at least for IP telephony traffic, while the need for other Layer 4-7 application awareness has not emerged. Instead of edge devices marking VOIP packets, the first switch/router in the path (or the IP-PBX) does so. Thereafter, VOIP traffic is frequently routed on separate paths from the mix of LAN data traffic, although they may be sharing the underlying network infrastructure.

This *de facto* IP telephony solution also incorporates the following elements:

- Better signaling protocols (SIP rather than H.323) and improved codecs, resulting in better-sounding voice.
- Upgraded L2 Ethernet switches that can prioritize SIP packets using 802.1 p/q.
- Specialized VOIP PBXs with separate pathways optimized for proprietary IP telephony.

This arguably has worked well for voice, but it won't necessarily help deliver other latency-sensitive applications, such as audio/video streaming or interactive conferencing. So far, the workaround for these applications has been memory buffering. That's fine for one-way viewing of movies, but it won't work for two-way videoconferencing with broadcast quality reception (let alone HDTV). For these kinds of applications, the solution will be one of the following:

- a.) adding video onto the voice PBX, although this will have to be done as another overlay so as not to risk reduced voice quality when the bigger video packets delay the smaller voice packets;
- b.) setting up a separate video server to host streaming and interactive applications
- c.) reintroducing the idea of a universal, high-

**New software in routers could accept security updates remotely, whenever they become available**

capacity, non-blocking switch/router that can read packets on the fly and provide differential quality of service.

In our view, option (c) is the simplest and best solution. Not only would it “fit” the mixed-application traffic flow demands of multiple types of enterprise networks, but it would also stimulate embryonic demand for multimode conferencing and streaming applications.

To do this, we’ll need a new generation of switch/routers. Current switch/routers can read and mark packet headers at 10/100 Mbps for L2 802.1p/q and L3 DiffServ QOS. Going forward, they’ll need to do this at 10 Gbps if they want to handle video, data and voice streams simultaneously. We have already noted the stresses of doing L3 routing at 10-Gbps speeds. Doing L4–7 processing will be even more difficult.

**Security Needs To Flex Its Muscles**

IPSec concepts have been incorporated in Cisco’s IOS, which is great for setting up VPNs as long as you use an all-Cisco network. Interoperability is something else.

More generally, however, the passing years have called into question the sufficiency of the IPSec idea. Adding extra bits in the header, as IPSec does to create secure tunnels, does nothing to prevent or mitigate denial-of-service attacks. The same is true for protection against spam, viruses or malware.

What would characterize the correct solution? Flexibility to deal with the ever-changing threats, as well as consistent execution. For example, if every PC had virus protection that was instantly updated whenever the user logged on, this would go a long way toward solving the problem. Unfortunately, like safe sex, everyone knows the right answer but few do it consistently.

It may be impossible to secure all the millions of PCs, but what if we focus instead on the considerably fewer and more readily accessed switch/routers? New software inside them could accept remotely transmitted programming instructions on the fly, including the latest spam and virus filters, in near-real time.

Network and IT shops have been ordered by upper management to more tightly secure their networks. This will take a level of agility that current devices don’t have, although the updating technology we just mentioned is available. The active network concepts funded by DARPA in the 1995–2000 timeframe could be commercialized to enable routers that can be programmed remotely, on a packet-by-packet basis, without penalty in price or performance.

**Wi-Fi, Meet Cell Phones, SIP And Bluetooth**

Since 1999, the number of wireless phones has grown to the point that they now outnumber wireline connections. That development alone means that we need to think seriously about linking the

**Decoding The Acronyms**

Acronyms	Definition
C, C++, C#	System software programming languages
DiffServ	Differentiated Services
FCS	Fibre Channel System
IP	Internet Protocol
IPSec	Internet Protocol Security
iSCSI	Small Computer System Interface over IP
LAN	Local Area Network
MPLS	Multi-Protocol Label Switching
MVNO	Mobile Virtual Network Operator
NAS	Network Attached Storage
NGN	Next Generation Networks
QOS	Quality of Service
PCI	Personal Computer Interconnect
SAN	Storage Area Network
SIP	Session Initiation Protocol
TCP	Transport Control Protocol
TDM	Time Division Multiplexing
TOE	Transport Control Protocol (TCP) Offload Engine
VOIP	Voice over Internet Protocol
VPN	Virtual Private Network
WAN	Wide Area Network
XML	Extensible Markup Language

two types of network more effectively. Wireless users need the same range of Internet-enabled low-cost services, including email, instant messaging and Web browsing, that are available today to wireline (and, soon, to Wi-Fi attached to wireline).

For example, consumers and SOHO business users clearly will benefit from the dual-mode cordless/cell phones that are coming to market. These will use Bluetooth or low-power Wi-Fi to link to home/office LANs, so that users can make and receive VOIP calls inside or outside the home or office.

Proprietary products for enterprise use also are close to market, but carriers might profit more from SIP phones with Bluetooth, and SIP-capable Wi-Fi switches that could route calls to the carriers’ wireline networks—assuming the phone is wireline-accessible—at much lower prices for the user. Cingular, Verizon, Sprint and T-Mobile probably won’t want to lose mobile traffic by making such connections, but AT&T—in its probable new manifestation as a mobile virtual network operator (MVNO) will have little to lose and much to gain by linking wireline and wireless. Again, we see the need for more powerful, intelligent and

flexible switch/routers that can handle the necessary protocol conversions and packet routing fast enough to preserve voice quality.

### Time To Streamline Storage

The continuing increase in storage capacity requirements over the past five years has been met predominantly by network attached storage (NAS) and storage area networks (SANs). In both cases, specialized servers connect to disk drive controllers that in turn connect to multiple disk drives. The main difference is that the SAN removes this traffic from the production data network. Demand has also grown for high-bandwidth storage applications such as mirroring and data backup at remote sites over the enterprise wide area network (WAN).

Storage applications require that data flow rapidly from the disk controller to the server, so that the data stream from the disk drives doesn't overwhelm memory buffers in the disk controller. Extending these functions to longer distances allows servers, controllers and drives to be in different buildings.

Fibre Channel systems (FCS) satisfy these requirements, and have been the *de facto* storage networking choice for many years, offering links of up to 10 km at speeds of 1 Gbps, 2 Gbps and 4 Gbps. Until recently, Ethernet's 10/100-Mbps links were too slow.

Fibre Channel has other protocol features designed to accommodate storage traffic's more rigorous timing, packet size and reliability requirements, which IP was not believed capable of handling. As a result, Fibre Channel SANs are connected to, but separate from the Ethernet/IP LANs that overtook enterprise networking.

However, new options have emerged in the past few years. We now have Gigabit and even 10-Gigabit Ethernet, the latter being faster than 4-Gbps FCS. We also have the Small Computer System Interface over IP (iSCSI), which can handle high-speed data storage transfers over Gigabit Ethernet LANs. So, if iSCSI takes hold, we could see the merging of networking and storage.

In the meantime, however, we have various niche products that have made a success in the SAN market, albeit at the cost of incomplete capabilities and incompatible standards. For example:

■ To access Fibre Channel from an Ethernet LAN, you need multiprotocol switches, such as those from McData, Brocade and Cisco that are specifically designed for this purpose. However, these switches only work at up to 1 Gbps. So at least for now, the transport on either side is faster than the switches in between. In addition, these multiprotocol switches are very expensive on a price-per-port basis compared to Ethernet switches.

■ Current SAN-to-LAN multiprotocol switches do nothing to facilitate seamless integration of Fibre Channel and iSCSI drives. In many

instances, accessing information from a SAN requires data that is stored on multiple disk drives. Current switches don't support such features as virtualization across disk arrays of differing protocol types (i.e., when some data blocks are on iSCSI storage and some blocks are on FCS storage). For all these reasons, iSCSI is often used for greenfield deployments, and kept isolated from Fibre Channel.

■ Products delivered over the last 18 months from vendors such as Network Appliance and others promise the convergence of NAS and SAN by allowing a single disk array to have both a block (SAN) and file (NAS) interface to common storage. However, a converged switch will be needed to make these products viable.

In the next generation, however, achieving this may not be possible in a single switch, because attaching a storage device to a NAS implies that the storage device controller has to handle all the communications functions of the network, which generates significant interrupt handling on processors. Some help is on the way, however, in the form of TCP/IP offload engines (TOEs), as well as TCP/IP proxy agents (for layers 5-7). The TOEs and the proxies could be located either in the switch or in the disk controller. They could help keep switch port prices down, by terminating and front-ending TCP/IP. (Note that Microsoft is re-implementing TCP/IP in Windows XP with offload hooks for TOE.)

Despite the venture capital that was lavished on storage in the bubble days, we don't have the multiprotocol switches today that we need. Many of the VC bets were made on software-based virtualized storage management (taking physical storage scattered over diverse servers, potentially at multiple nodes in a network, and making this a logical or virtual storage server). In contrast, apart from beefing up Fibre Channel from 2 to 4 Gbps, there's been relatively little hardware activity. In our opinion, that needs to change.


The partial products on offer today—even with the TOE additions—point up the need for a more comprehensive solution: A multiprotocol switch that processes TCP/IP, translates Ethernet, iSCSI and Fibre Channel seamlessly, and does so at up to 10-Gbps rates. Such a box would truly streamline storage, backup, mirroring and network management.

### Extend The Power Of Grid Computing

At its most basic level, grid computing lets users access multiple computers and storage disks transparently, without concern as to where the resources are located (source: [www.genomicglossaries.com](http://www.genomicglossaries.com)). This idea is gaining appeal, in large part because of its successful application by companies like Google.

Not only did Google spend a fraction of what other search engine companies spent on its hardware (according to Google's website, Google had

**With iSCSI and 10-Gigabit Ethernet, Fibre Channel's days dominating SANs may be numbered**



**Grid computing  
might stimulate  
the need for  
non-blocking,  
high-capacity  
Ethernet switches**

more than 100,000 servers some months ago) but the company also popularized the grid computing notion of algorithms that execute simultaneously on multiple processors, with a goal of keeping each server in each node active executing software. Grid computing also is a key underpinning of IBM On Demand, HP Adaptive Enterprise and the University of Virginia's parallel-processing supercomputer, among others.

To set up a grid computing network, you put a number of microprocessor cards on a shelf, with a PCI bus connecting the shelves. On each shelf, you also put in a switching board that links the shelf to a central switch. This switch, in turn, accesses a NAS, where the data that the algorithms retrieve are stored.

As with SANs, people have tried to develop specific high speed protocols to replace the PCI bus linking together these microprocessor shelves, with Infiniband posited as the grid computing analogue to Fibre Channel. Unfortunately (and unlike the case with Fibre Channel), Infiniband hasn't gained much market traction, and 1-Gbps/10-Gbps Ethernet appears to be preempting the need for a distinct grid computing protocol (Google uses both 1-GigE and 10-GigE).

If we connect the microprocessor arrays to an Ethernet fabric that acts not only as the grid communication fabric, but also connects the grid to the LAN and SAN, we could use a Brocade switch or a newer multiprotocol switch on the SAN side, and Veritas software to manage the storage. But we would still need more robust non-blocking switches to provide a converged fabric that connects LAN and SAN.

In 1999, people used to talk about the need for terabit and petabit switches, largely for carriers at the network core. VCs funded a number of high-capacity switch startups, most of which failed to deliver, including Argon, NetCore, Nexabit and others. In contrast, as grid computing gains traction, we are going to need very high capacity, non-blocking Ethernet switches in the enterprise.

The reason for this is simple. If an important objective of grid computing is to have each microprocessor and each disk drive operate at high capacity utilization, this means that for any level of microprocessing power and disk drive capacity, there will be more switching taking place in grid systems than in non-grid systems and with it, more contention. Ethernet is an excellent bus arbitration algorithm for backplanes on switches (arbitration time is the time for signals to propagate across the backplane). So, for example, if we have a 256-PC microprocessor array, each operating at 1-Gbps bus speed and each transmitting/receiving most of the time, we will need a switch with 256 Gbps of non-blocking capacity. If we go to 10-Gbps buses, we'll need 2.6 terabits per second of capacity. Today, we don't have such high-capacity, non-blocking switches at low enough cost for use in the enterprise.

### **Fat XML Files Will Need Bigger Pipes**

XML's ease of use, programmability and extensibility are making it the application language of choice for Internet-based applications, including Web services. The trade-off, however, is that XML generates files 10 times the size of comparable C, C++, or C# files. No problem in the lightly-loaded Gigabit networks of the application developer labs, but downloading these 10MB, and even 100MB XML files over today's mostly 10/100-Mbps enterprise networks will swamp most of them, leading to major congestion. This is not widely discussed in networking literature. Instead, what we see are anecdotes about XML file transfers, with eyes averted whenever the bandwidth requirements come up.

Yet if XML really is the development language of choice over time, and if we're going to see lots more XML traffic, we need to think about ways to optimize the network for this traffic. If we were starting with a clean sheet of paper, we would want to (1) allow something like "jumbo frames" for XML, whose packets can vary in size by orders of magnitude, and (2) reduce per-packet overhead and processing.

We would then have a network in which we could transmit native XML, rather than XML over Ethernet. The benefit would be substantially reduced switching time to handle large XML packets. If there is a cost, it is that an XML switch would need substantially more buffering memory than an Ethernet switch of equivalent throughput.

A number of switch/router startups are being funded now to address this need, including Flamenco, SlamDunk, and Grand Central Communications. These may be appropriate for very large enterprises with substantial XML traffic between servers and NAS, and where native XML switches could make XML grid computing applications more efficient. However, in most of the rest of the world, Ethernet and IP are ubiquitous, so a substantial portion of XML traffic will likely need to travel over IP/Ethernet networks.

Again we see the need for a general purpose, high capacity switch/router that can gracefully handle multiple protocols, at wire speeds, on the fly: XML, TCP/IP, Ethernet, MPLS, ATM, FCS, iSCSI, SONET/SDH. Why not build such a box — especially since advances in silicon mean that it will cost no more to do so?

### **Conclusion: A Next-Gen Architecture For Today**

Over the past 10 years, point solutions have won out in the marketplace over the muscular, multi-purpose devices — sometimes deprecated as "god boxes" — which we are again championing here. For example, Brocade switches do a great job translating Fibre Channel and Ethernet, and IP-PBXs satisfy their customers by segregating VOIP traffic. Those approaches are good near-term solutions.

The problem with point solutions is that, if they

are successful, they have a tendency to stack up—as have the firewalls, NAT boxes, VPN concentrators, application proxies and bandwidth managers that now crowd the borders of most enterprise WANs. Layering lots of point-solution boxes throughout your network takes more space, power and cabling than an integrated solution. In principle, then, the “god box” solution makes better sense because it reduces these costs.

God boxes got a bad rap in the late '90s, when they promised much but delivered little. Few people bought them, because they did not work as advertised and they were too expensive. Fortunately, technical and price-performance advancements in silicon integrated circuitry have solved these problems.

It is our opinion that the industry is set to repeat a cycle similar to that which Cisco capitalized upon in the late '80s. At that time, companies had many networking protocols: DECNet, IPX, SNA, Appletalk and others. Cisco's multiprotocol router successfully handled all of them.

Since then, of course, Ethernet and IP have “won,” but in the past few years VOIP, wireless, storage and grid computing have splintered off to develop their own switch/routing solutions. This time, we believe talk of “god boxes” won't be frivolous—it will be necessary.

A lot depends on which companies develop market traction for the various point solutions in the technology areas we've been discussing. While it's difficult to say how this will play out, there is clearly a strong future for agile, multiprotocol, scalable switch/routers in supporting the evolving needs of the enterprise network□

#### Companies Mentioned In This Article

Brocade Systems ([www.brocade.com/](http://www.brocade.com/))  
Cisco ([www.cisco.com](http://www.cisco.com))  
EMC ([www.emc.com](http://www.emc.com))  
Flamenco ([www.flamenconetworks.com](http://www.flamenconetworks.com))  
Google ([www.google.com/](http://www.google.com/))  
Grand Central ([www.grandcentral.com](http://www.grandcentral.com))  
Hewlett Packard ([www.hp.com/](http://www.hp.com/))  
IBM ([www.ibm.com/](http://www.ibm.com/))  
Intel ([www.intel.com/](http://www.intel.com/))  
McData ([www.mcdata.com/](http://www.mcdata.com/))  
Microsoft ([www.microsoft.com/](http://www.microsoft.com/))  
Network Appliance ([www.netapp.com/](http://www.netapp.com/))  
SlamDunk ([www.slamdunknetworks.com](http://www.slamdunknetworks.com))  
Storage Tek ([www.storagetek.com/](http://www.storagetek.com/))  
Veritas ([www.veritas.com/](http://www.veritas.com/))



**God boxes got a bad rap in the late '90s—but they might be needed in the years to come**