

Blade-level switching: A key to upgradeable Layer 4-7 services

By Uri Cummings
July 7, 2005

Many enterprises used to upgrade their networks solely for the purpose of meeting their end users' increased demand for bandwidth. Initially, these upgrades meant transitioning from bus-based shared networks to switched Ethernet networks. The next step was migrating the Ethernet networks to switched fast Ethernet networks.

While they still need to meet pressing bandwidth needs, today, enterprises entertaining the prospect of transitioning from fast Ethernet to gigabit Ethernet (GbE) networks have additional reasons to upgrade. This time, they are building network infrastructures that can deliver increasingly sophisticated network services and intelligence to end-users. Generally, this means venturing into the upper layers of the protocol stack as they make those network upgrades.

This article explores the evolution of the networking system designs used to accommodate these trends with a focus on board-level interconnect. Board-level interconnect represents the last bottleneck in board design because traditional daisy-chain and bus-based interconnects limit flexibility and can have high latency and throughput that is impacted by bus contention and speed limitations. As a result, switch-based interconnect architectures are becoming increasingly popular in gigabit-speed designs used to support the improved performance and data flow complexity.

New data flows drive change

The drivers increasing demand for network services are new types of data flows in the network. These new data flows include:

- near-real-time data for financial database and enterprise software applications
- real-time network data flows for voice over IP (VoIP)
- multicast video
- other streaming media.

The addition of these data flows has precipitated enterprise-wide reliance on the network. This reliance, in turn, has been accompanied by an increase in the need for network reliability and data security.

To support these new and growing requirements, it is important for a networking device to look more deeply into a packet, oftentimes beyond Layer 3 routing information. Information in Layers 4-7 can help a switch support these new data streams.

For example, by parsing a packet and reading transmission control protocol (TCP) data at Layer 4, a switch can recognize VoIP data packets and apply the appropriate quality of service (QoS) mechanisms to that packet. At



the application layer, Layer 7, a switch can facilitate server load balancing. Requests that come into a Web site's virtual IP address can be distributed according to their type. For example, e-commerce sessions can be directed to one server, while video download sessions can be directed to another server.

Interconnect: The last bottleneck

Board-level architecture for these systems has had to evolve to support new services. This has meant increased use of NPUs, security co-processors and high-speed interconnect protocols like SPI-4.2 for the data path, and PCI Express for control. However, the impact on board-level interconnect extends beyond higher speed protocols.

When services were limited to Layer 3 and below, the data path through the network could be a linear daisy chain. This data plane could be optimized for wire-speed performance with fast ASICs providing the services. When an exception occurred, such as an SNMP trap or a routing table update, packets involved could be taken out of the data plane and processed off line before being re-inserted into the data flow.

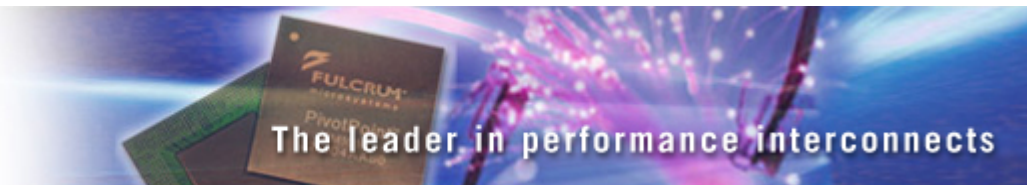
With an increased level of Layer 4-7 functions, the data flow associated with today's services has become more complex (see Figure 1). A larger number of packets now need additional processing " at wire speed " and some must utilize several resources before they can be forwarded by the switch to their final destination. An encrypted packet, for example, arrives at an NPU only to be sent off to a security co-processor before coming back to that NPU for processing and forwarding to its end destination. In short, the straight, linear data plane is yielding to a complex traffic pattern that can be different for each packet that passes through the system.

A new interconnect

An emerging trend in this design environment is the implementation of a switched interconnect. This trend is demonstrated by the evolution of several popular bus standards to support switched interconnects. The standards bodies and promoters of HyperTransport, RapidIO and PCI Express have positioned each of these standards to evolve to switched-bus designs. In addition, commercial crossbar switch chips are available from a growing number of vendors. These chips are full switch implementations with reprogrammability and flow control capabilities. The first chips to emerge support the SPI-4.2 protocol, and other protocols are under development.

Making the transition to a switched network requires equipment designers to learn new terminology and be alert to a few design considerations. A switch's performance is determined by its capacity, which is a combination of the chip's port count, the speed of those ports and the throughput of the switch fabric. On an SPI-4.2 switch, for example, each port operates at up to 16 Gbit/s. If the chip has two ports, the switch fabric must support at least 32 Gbit/s in order to keep data from being blocked. Often switch fabrics are designed to support a multiple of the chip's total port speed. This is called overspeed.

Switch latency is the time it takes a data packet to traverse the switch. Latency is as critical as throughput in selecting the proper switch. There are two kinds of switching paradigms, and each can have a dramatic impact on latency. In cut-through switching, a packet's destination is read and the packet is forwarded through the



switch before the switch receives the entire packet. In store-and-forward switching, by contrast, the entire packet must enter the switch before it can be forwarded. Generally, a cut-through switch will offer the lowest latency, especially on designs that anticipate large packets.

If a switch with significant internal overspeed does not also offer low-latency transport and low-latency flow control, throughput problems are likely to occur. Bursty traffic can be translated into sporadic "sawtooth" performance due to an oscillation between congestion, delayed flow control, and drained buffers. Designers need to tune the size of their buffers to the speed and latency of the switch chip. Larger buffers will result from either an under-speed product, in which the switch fabric processes less than the aggregate of the link speeds, or a product with high latency.

Flow control also is important to system performance, and it is significantly impacted by latency. Flow control through a switch must be broken down to the link level so that at each stage " ingress, in transit, and egress " there is the opportunity to report congestion and cause the sender to stop transmitting sooner. The more tightly coupled the flow control and lower the latency of the flow control path, the more efficiently the overall system will operate.

The interface ports on a switch should be channelized to the extent required to support the application. In the case of SPI-4.2, this can be as high as 256 channels, although far fewer channels are required for most applications. These channels, sometimes called ports, may be used to transport separated traffic flows. They have distinct hardware resources, buffers and flow control mechanisms, but share a common physical interface. The ability to map any channel on any interface to any other channel on any other interface is also important to allow complex data flows through multiple devices.

A key issue is head of line (HOL) blocking where a packet is blocked at the ingress port by congestion in the switch and that keeps other packets, whose data paths are not affected by the congestion, from progressing. HOL blocking will affect the actual throughput of the switch regardless of the switch's capacity. Port channelization and per-channel flow control are important features within a chip because they can be used to overcome HOL blocking.

New possibilities

One of the requirements of many new networking designs is the need for data to flow back (see Figure 2). Packets often must be processed multiple times by multiple resources before they can be forwarded. For example, an encrypted packet arriving at an NPU for classification must be sent to a security co-processor for decryption before it flows back to the NPU for final processing.

To accomplish this without a switched interconnect would mean putting a security co-processor in the data plane with the NPU and requiring all packets to filter through it just to accommodate the few packets that are encrypted. In this design, the security processor becomes a bottleneck if it does not operate at wire speed.



An extension of this switched design adds a "mid-plane" to the design, in which security, SNMP, content management and other resources reside, with the switch feeding packets to each resource as necessary. This optimizes the performance of the data plane, while allowing the design to offer Layer 4-7 features.

The concept of a switched interconnect also is central to some of the design flexibility benefits built into the new Advanced Telecom Computing Architecture. The ATCA is the first standard created for the system design of central office-based networking equipment.

Built into the ATCA specification is support for advanced mezzanine cards (AMCs), which can be inserted into a carrier card for added functionality, physical interface flexibility, or added computing or storage networking capability. With a switch interconnect on the carrier card, these AMCs can be connected and mixed and matched as needed. Taken to the extreme, A highly-modular system composed on best-of-breed AMC building blocks could lead to a universal access line card that can be populated with whatever mix of Ethernet, DSL, Wi-Fi or other connectivity standard that is needed for the application.

Conclusion

A switched interconnect can offer a new level of flexibility to networking designs. With more vendors entering the market and improved switching support built into interconnect protocols, engineers will have more choices in how they can reshape their architectures to benefit from switching.