This story appeared on Network World at
http://www.networkworld.com/news/tech/2012/101112-next-gen-infrastructure-263297.html

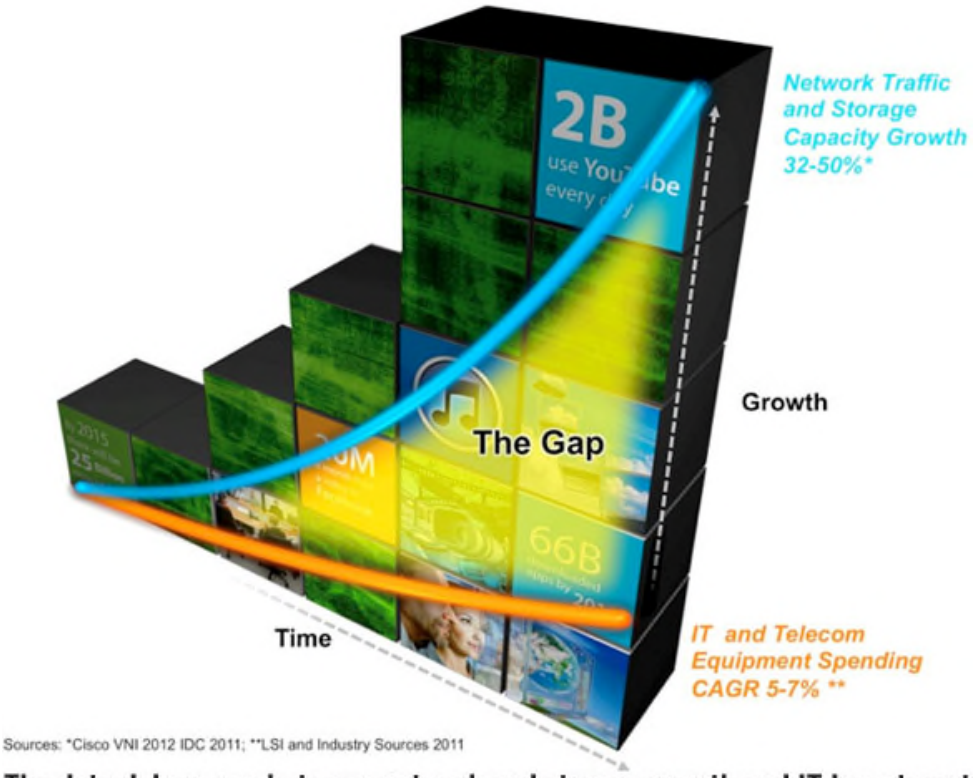
Why next-generation infrastructures need smarter silicon

By Jim Anderson, senior VP and general manager, Networking Solutions Group at LSI
October 11, 2012

Given the explosive growth in data traffic, Moore's Law is not enough to keep pace with demand for higher network speeds. A smarter silicon and software approach is needed.

Among the best ways to accelerate the performance of mobile and data center networks is to combine general-purpose processors with smart silicon accelerator engines that significantly streamline the way bits are prioritized and moved to optimize network performance and cloud-based services.

One of the fundamental challenges facing the industry is the data deluge gap -- the disparity between the 30% to 50% annual growth in network and storage capacity requirements and the 5% to 7% annual increase in IT budgets. The growing adoption of cloud-based services and soaring generation and consumption of data storage are driving exponential growth in the volume of data crossing the network to and from the cloud. With the growth in data traffic far outstripping the infrastructure build-out required to support it, network managers are under pressure to find smarter ways to improve performance.



The data deluge gap between network and storage growth and IT investment is wide and getting wider.

Cloud data center networks were built with existing technologies and have thus far succeeded in improving performance through brute force -- adding more hardware such as servers, switches, processor cores and memory. This approach, however, is costly and unsustainable, increasing hardware costs along with floor space, cooling and power requirements, and falls well short of solving the problem of network latency.

Adding intelligence in the form of smarter silicon streamlines processing of data packets traversing mobile and data center networks. In particular, smart silicon enables next-generation networks to understand the criticality of data, then manipulate, prioritize and route it in ways that reduce overall traffic and accelerates the delivery of important digital information, such as real-time data for voice and video, on time.

Smarter networks

General-purpose processors, which increasingly feature multiple cores, pervade network infrastructures. These processors drive switches and routers, firewalls and load-balancers, WAN accelerators and VPN gateways. None of these systems is fast enough, however, to keep pace with the data deluge on its own, and for a basic reason: general-purpose processors are designed purely for compute-centric, server-class workloads and are not optimized for handling the unique network-centric workloads in current and next-generation infrastructures.

Smart silicon, however, can accelerate throughput for real-time workloads, such as high-performance packet processing, while ensuring deterministic performance over changing traffic demands.

Smart silicon typically features multiple cores of general-purpose processors complemented by multiple acceleration engines for common networking functions, such as packet classification with deep packet inspection, security processing and traffic management. Some of these acceleration engines are powerful enough to completely offload specialized packet processing tasks from general-purpose processors, making it possible to perform switching, routing and other networking functions entirely in fast path accelerators to vastly improve overall network performance. Offloading compute-intensive workloads to acceleration engines that are optimized for a particular workload can also deliver a significant performance-per-watt advantage over purely general-purpose processors.

Customized smart silicon can be a great option for a network equipment vendor wanting to carve out a unique competitive advantage by integrating its own optimizations. For example, a vendor's proprietary, differentiating intellectual property can be integrated into silicon to provide advantages over general-purpose processors, including for optimized baseband processing, deep packet inspection and traffic management. This level of integration requires close collaboration between network equipment and semiconductor vendors.

Tomorrow's data center network will need to be both faster and flatter, and therefore, smarter than ever. One of the key challenges to overcome in virtualized mega data centers is control plane scalability. To enable cloud-scale data centers, the control plane needs to scale either up or out. In the traditional scale-up approach, additional or more powerful compute engines, acceleration engines or both are deployed to help scale up networking control plane performance.

In emerging scale-out architectures like software-defined networking (SDN), the control plane is separated from the data plane, and then typically executed on standard servers. In both scale-up and scale-out architectures, intelligent multicore communications processors that combine general-purpose processors with specialized hardware acceleration engines can dramatically improve control plane performance. Some functions, such as packet processing and traffic management, often can be offloaded to line cards equipped with these purpose-built communications processors.

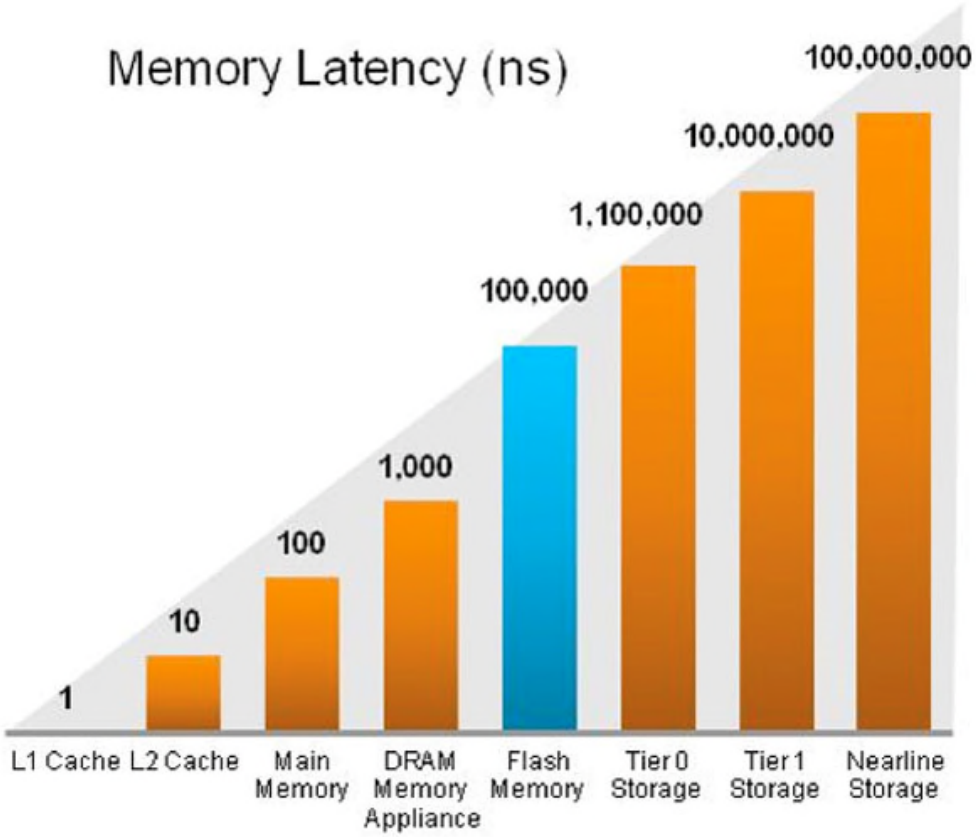
While the efficacy of distributing the control and data planes remains an open question, it's clear that SDN will need smart silicon to deliver on its promise of scalable performance.

Smarter storage

Smarter silicon in storage can also help close the data deluge gap. The storage I/O choke point is rooted in the mechanics of traditional hard disk drive (HDD) platters and actuator arms and their speed limits in transferring data from the disk media, as evidenced in the difference of five orders of magnitude in I/O latency between memory (at 100 nanoseconds) and Tier 1 HDDs (at 10 milliseconds).

Another limitation is the amount of memory that can be supported in traditional caching systems (measured in gigabytes), which is a small fraction of the capacity of a single disk drive (measured in terabytes). Both offer little room for performance improvements beyond increasing the gigabytes of Dynamic RAM (DRAM) in caching appliances or adding more of today's fast-spinning HDDs.

Solid state storage in the form of NAND flash memory, on the other hand, is particularly effective in bridging this significant bottleneck, delivering high-speed I/O similar to memory at capacities on a par with HDDs. For its part, smart silicon delivers sophisticated wear-leveling, garbage collection and unique data reduction techniques to improve flash memory endurance and enhanced error correction algorithms for RAID-like data protection. Flash memory helps bridge both the capacity and latency gap between DRAM caching and HDDs, as shown in the chart.



Flash memory fills the gap in both latency and capacity between Dynamic RAM in a cache appliance and fast-spinning hard disk drives.

Solid state memory typically delivers the highest performance gains when the flash cache acceleration card is placed directly in the server on the PCI Express (PCIe) bus. Embedded or host-based intelligent caching software is used to place "hot data" in the flash memory, where data can be accessed in 20 microseconds -- 140 times faster than with a Tier 1 HDD, at 2,800 microseconds. Some of these cards support multiple terabytes of solid state storage, and a new class of solution now also offers both internal flash and Serial-Attached SCSI (SAS) interfaces to combine high-performance solid state and RAID HDD storage. A PCIe-based flash acceleration card can improve database application-level performance by five to10 times in DAS and SAN environments.

Smart silicon is at the heart of all of these solutions. So without the deep inside view of the semiconductor vendors, the system vendors would have no hope of ever closing the data deluge gap.