New I/O Technologies Seek to End Bottlenecks

George Lawton

New processors, hard drives, memory chips, networking standards, and other technologies are pushing the limits of computer capabilities. However, the improved performance these developments promise for PCs, servers, and even embedded systems is being choked at a bottleneck caused by input/output technology, which has failed to keep pace.

For example, said Peter Glaskowsky, analyst with MicroDesign Resources, a semiconductor-industry research firm, users cannot get the full performance of a Gigabit Ethernet network card for some peripheral component interconnect (PCI) products because the card runs too fast for the bus. PCI, the mainstay of PCs and embedded systems, is almost a decade old and hasn't undergone substantive changes.

To cope with this problem, vendor groups are releasing new I/O platforms such as PCI-X, InfiniBand, and RapidIO. The new technologies promise faster performance, reduced latency, reduced CPU computational load, and more efficient and reliable systems.

However, as is the case with many new technologies, there are concerns. For example, industry observers speculate that the new I/O technologies could become so popular, companies may use them inappropriately, thereby reducing their effectiveness.

In addition, some users are worried that the different technologies are incompatible or that bridges between technologies could add programming complexity.

DRIVING THE NEW I/O

The Industry Standard Architecture, which IBM released about 20 years ago to work with its early PCs, was the first I/O bus to gain widespread support.

ISA is still widely used on PCs today, along with PCI. However, ISA limits communications with peripherals to a 16-bit bus running at eight to 10 MHz with a maximum throughput of 160 Mbits per second. EISA (extended ISA), which an industry consortium developed, doubled the transfer rate with a 32-bit bus.

Intel developed PCI in the early 1990s with a 32-bit bus and a faster data rate of 33 MHz.

However, PCI has important limitations. For example, it was not designed to work with the networking technologies that are so important to computing today.

PCI can connect devices only within a few feet of a motherboard and offers up to only six expansion slots on a peripheral card. Also, all peripherals are electrically connected on the same bus, so a fault in one card can bring down the entire system. And because peripherals share the bus's resources, the data rate is limited while latency can increase.

In 1996, several major systems vendors such as Hewlett-Packard, IBM, and Sun Microsystems tried to go beyond PCI with I2O (intelligent input/output), which was supposed to provide high-speed links between machines. However, the I2O Special Interest Group could not agree on an approach and disbanded.

NEW I/O TECHNOLOGIES

Each new I/O technology targets a different niche market. PCI-X (PCI extended), like its predecessor, will specialize in connecting multiple components within the same box. InfiniBand, a serial technology geared toward fast external I/O, will link multiple servers with one another and with high-speed networking interfaces. RapidIO, a parallel technology, focuses on embedded systems such as telephone switches and routers.

The three technologies can work with one another via hardware-software bridges.

PCI-X

Vendors are just beginning to release servers and peripheral cards based on PCI-X, a specification developed by major PC vendors such as Compaq Computer, HP, IBM, and Intel to improve PCI's price-performance ratio.

Roger Tipley, a principal member of Compaq's technical staff and the PCI Special Interest Group's president, said the PCI-SIG (http://www.pcisig.com) would push to get PCI-X deployed in all PCs starting in 2002.

PCI-X differs from PCI primarily in that it implements register-to-register signaling. RRS more accurately resolves signals, so the same size bus can have higher clock and data rates. In essence, RRS can double the speed at which the I/O reads data (although the write performance remains about the same) without increasing the cost or size of PCI components.

With PCI, one 64-bit bus runs at 66 MHz and additional buses transmit either 32 bits at 66 MHz or 64 bits at 33 MHz, allowing a top data rate of 532 Mbytes per second. With PCI-X, one 64-bit bus runs at 133 MHz and the others transmit

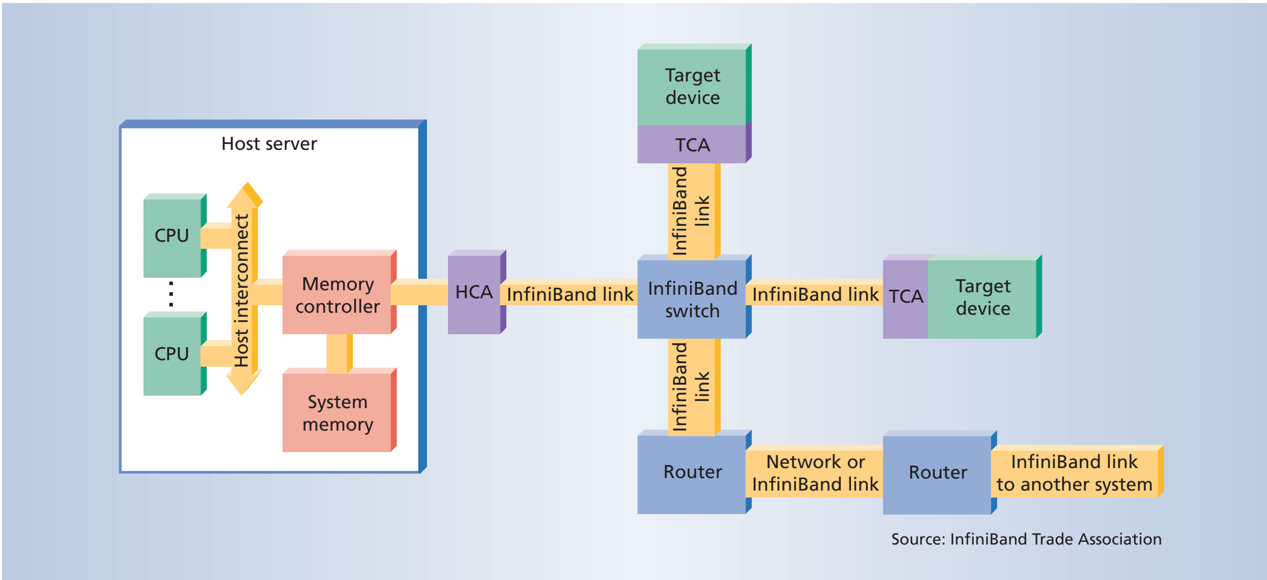

Figure 1. With InfiniBand switched-fabric I/O technology, a host server connects to a host channel adapter (HCA), while each target server or peripheral device connects to a target channel adapter (TCA). The adapters, which act as intelligent interfaces between devices, communicate through an InfiniBand switch.

at 66 MHz, allowing a maximum data rate of 1.06 Gbytes per second.

PCI-X also limits the amount of time peripherals must wait to access the bus channel, thereby reducing latency.

PCI-X buses and PCI peripheral cards are hardware and software compatible but not always electrically compatible because some PCI versions operate at different voltages than PCI-X. However, this will be resolved by PCI version 2.3.

In the future, Tipley said, proponents will try to double and then quadruple PCI's data rate to 2 and then 4 Gbytes per second, to better work with faster networking technologies.

Initially, PCI-X will be more expensive than PCI, but with time and mass production, the cost will be about the same, said Jonathan Eunice, principal server analyst at Illuminata, a consultancy and market research firm.

InfiniBand

InfiniBand is an important new I/O technology for the high-end server market. The technology was formed by a merger of two competing projects: Future I/O (backed by such companies as Cisco Systems, Compaq, HP, and IBM) and Next Generation I/O (formed by Intel and supported by such companies as Dell Computer and Sun Microsystems).

The InfiniBand Trade Association (http://www.infinibandta.com) released the first specification earlier this year, and more than 45 vendors plan to begin releasing enabled products, such as networking cards for servers and storage systems, later this year.

InfiniBand is a switch-based architecture for connecting up to 64,000 CPUs, storage systems, and other components at data rates of 2.5, 10, or 30 Gbits per second, depending on the cabling between communicating devices.

The technology operates across a range of distances depending on the transmission medium, from 17 meters over copper wire to 10 kilometers over single-mode optical fiber.

TCP/IP. The InfiniBand architecture is based on IPv6 addressing. Each node is given an IPv6 address. Because IPv6 uses a 128-bit address field, InfiniBand can link a large number of devices.

Allyson Klein, industry marketing manager for Intel's InfiniBand initiative, said, "With InfiniBand, TCP/IP will be pushed to the edge of the data center."

InfiniBand would enable development of an architecture capable of higher speed Internet connections, said Illuminata's Eunice. Edge servers would translate TCP/IP packets into InfiniBand packets, which entail less overhead for reliability and interoperability, and route them to the most appropriate server, he explained.

In addition, the edge server would handle some of the processing load otherwise run by the CPU. This reduces the CPU's burden and permits faster processing, because the CPU must perform other tasks in addition to bus-related computing.

Switched fabric. In an InfiniBand system, each host server connects to a host channel adapter (HCA), while each target server, storage system, router or other peripheral device connects to a target channel adapter (TCA), as Figure 1 shows. The adapters function like intelligent network interface cards.

The server and devices then communicate, through their adapters, via an InfiniBand switch located between the machines. The switch's intelligence manages the linkage without interrupting the system's operation.

In addition, the adapters can exchange information that ensures desired security or quality-of-service (QoS) levels. They can also determine when the bus is having a connection problem and look for alternative data routing. PCI cannot do this.

The InfiniBand switch temporarily opens up channels between the processor and the devices with which it is communicating. The devices thus don't have to share a channel's bandwidth, as is the case with a bus-based design like PCI, which forces devices to arbitrate for access to the processor.

The ability of devices to connect directly to the bus, without waiting, also reduces data-transfer latency.

Users can attach additional devices simply by connecting their TCAs to the InfiniBand switch via copper wires or fiber cables. The technology does not use expansion slots, as is the case with PCI.

And because InfiniBand uses a switching fabric in which power is provided independently to each component, a user could add capacity without bringing down the system.

Intel's Klein said, "We created this technology with data centers specifically in mind. It [eliminates many] problems, especially in the Internet data realm where people are adding 10 or 20 servers a day."

However, InfiniBand requires a lot of processing to handle the technology's protocol stack. To manage this processing, Illuminata's Eunice said, InfiniBand should be implemented with a coprocessor.

RapidIO

Motorola presented the proposed standard for the technology to the RapidIO Trade Association (http://www.rapidio.org) for debate and modification before it was released in March, according to Richard Jaenicke, director of product marketing at Mercury Computer Systems, a vendor of high-performance digital-signal and image-processing systems, including RapidIO products. The association includes about 40 vendors, including Cisco Systems, IBM, and Nortel Networks. RapidIO 1.1 is now available on the Web at http://www.rapidIO.org/specs.html.

The technology is being used to improve embedded systems' performance with a switched architecture that, like InfiniBand, increases the data rate and reduces latency of each component.

RapidIO also provides a control-plane interface, which improves management and the ability to implement applications requiring high QoS levels.

InfiniBand links up to 64,000 servers, storage systems, and other components.

RapidIO transfers data at rates between 8 and 64 Gbits per second. It is best suited for connecting up to 256 devices with 8-bit addressing or up to 65,536 devices with 16-bit addressing over short distances.

Embedded-systems developers have traditionally used a PCI bus. Unlike PCI, which uses transistor-transistor-logic signaling, RapidIO uses low-voltage-differential signaling. LVDS communicates using the voltage difference between two wires, rather than the voltage level in a single wire, as is the case with TTL. LVDS experiences less electrical interference, so the systems can more accurately decipher signals.

Also, with RapidIO, all processing is done in hardware, on a CPU, field-programmable gate array (FPGA), or application-specific integrated circuit, depending on the application. This eliminates the need to write I/O software, run the software on a CPU (which slows the processor down), or provide a dedicated processor to run it (which increases costs).

In fact, an entire RapidIO interface can be implemented on only 30,000 gates of a 1-million-gate FPGA. By contrast, an InfiniBand implementation requires one or more of the chips.

Reducing component costs and size is an important consideration for embedded systems, which must be small and inexpensive.

Jaenicke said RapidIO will be most in demand for use in networking equipment applications such as telephone switches and routers because of its high bandwidth and reliability.

CONCERNS

The new I/O technologies could become so popular that they are used where they are not most effective, thereby generating consumer dissatisfaction, according to David Murray, vice president at TransDimension, an embedded chipset vendor.

Meanwhile, adoption of the technologies could be slow if there are problems with the initial implementations, said Illuminata's Eunice.

Finally, Eunice added, none of the improvements are addressing long-distance I/O connections. For example, he said, InfiniBand works well within a data center, but large organizations need a high-speed I/O technology to link data centers over a wide area.

I/O'S FUTURE

Some I/O improvements will lower the overall cost of high-speed communication between systems, thereby encouraging increased use of the new technologies.

Meanwhile, Eunice said, InfiniBand and RapidIO will upgrade performance and efficiency so much that organizations could improve data-center layout by squeezing up to 10 times the equipment, including servers and telephone switches, into a given area.

This, in turn, would let third-party data-center-hosting services such as Exodus Communications manage larger systems with the same number of employees.

The new I/O technologies will open up bottlenecks within and between systems and let IT managers run server systems more reliably and easily, Eunice said.

"People figured out how to do this in the 1970s for mainframes," he said, "but they are just now getting to the servers." ✷

George Lawton is a freelance technology writer based in Brisbane, California. Contact him at glawton@glawton.com.

Editor: Lee Garber, *Computer*, 10662 Los Vaqueros Circle, PO Box 3014, Los Alamitos, CA 90720-1314; l.garber@computer.org