Guest Editors' Introduction: Design Challenges for High-Performance Network Interfaces

Andrew A. Chien, University of California, San Diego
Mark D. Hill, University of Wisconsin-Madison
Shubhendu S. Mukherjee, Compaq Computer Corporation

With the advent of distributed computing, the Internet, the continued increases in processor clock rates, and complete computer systems on a single chip, computing is increasingly concerned with efficient movement of data–through the wires within a machine, across a system-area network (SAN) within a machine room, and across local-area or wide-area networks (LANs and WANs). Thus, an increasingly critical issue in the design of computer systems is how to achieve efficient communication and particularly, the design of network interfaces – the topic of this special issue. A network interface is a device that allows a computer to communicate with a network. Figure 1 shows a conventional network interface attached to an Input/Output (I/O) bus. Network interfaces can also be attached to memory controllers or even directly to the processor datapath.

Network interface design has a crucial impact on communication efficiency, determining the cost of initiating and responding to communication actions, providing application isolation, and the data movement required to achieve the communication. The interaction overhead between a processor and a network interface is exacerbated by increasing operation rates produced by microprocessors with gigahertz clocks and networks with gigabytes-per-second bandwidth.

Two important components of communication are:

- network software that manages the network and implements communication protocols and
- network interface hardware that moves data, provides protection, and generates communication events.

However, these components manifest themselves in different forms, depending on the application, system context, and cost-performance requirements. Three major classes of network connectivity are:

- *Workstations or PCs connected by a local area network (LAN).* Traditionally, LANs (and WANs) have provided unreliable delivery, and thus the computers use network software, such as TCP/IP protocol stacks, to ensure reliable delivery. The cost of such protocol stacks is significant. Consequently, network software is the key research target in this area. Performance optimizations include reducing the code path of protocol stacks and optimizing data movement between host memory and the network interface.

- *Workstations or PCs connected by a system area network (SAN).* SANs (e.g. Myricom's Myrinet and Compaq's Servernet) will deliver bandwidths of 10 Gbps and latencies of tens of nanoseconds---two to four orders of magnitude better than current LANs---to hosts in close proximity (e.g., 100 meters). These high performance levels and generally reliable delivery inspire the use of lightweight protocols (i.e. Active Messages or Fast Messages) and innovative protection and notification structures. SAN-based computing provides a promising avenue for building large-scale systems (in computing, memory, and storage) using low-cost building blocks.

- *Tightly-coupled Massively-Parallel Processors (MPPs) connected with a custom network*. MPPs are tightly integrated systems with the highest performance levels for communication and the deepest integration of such communication into the computing complex. This requirement is driven by fine-grain parallel applications that demand extremely low-latency and high bandwidth communication.

> Primary research objectives in this area include: reduction in the end-to-end latency and overhead of interaction between the processor and network interface. Performance optimizations include tighter integration of the network interface hardware with computing elements, such as the processor core.

Our special issue brings together four papers that address the key issues in network interface design across a wide range of cost-performance. The first paper—*Efficient High-Speed Data Paths for IP Forwarding using Host Based Routers*—by Walton et al. examines the use of commodity workstations, and enhanced network interface cards to route internet protocol (IP) packets. The authors show that routing performance can be boosted by direct transfers of the packet payload between the source and destination network interfaces, instead of staging it through the host memory.

The second paper—*Design Issues for User-Level Nework Interface Protocols on Myrinet*—by Bhoedjang et al. is a tutorial on network software design for SANs. This paper provides a broad perspective on the wide range of high speed protocols and interfaces built on Myrinet's network hardware as well as insights into critical design issues, such as data transfer, address translation, protection, and control transfer.

The third paper—*A Review of Message-Based User-Level Network Interfaces*—by von Eicken and Vogels examines how several university research projects on SAN NIs helped shape the industry-standard NI specification called Virtual Interface Architecture (VIA). The VIA standard is widely supported (several hundred companies) and is an emerging standard for cluster or system-area networks being jointly promoted by Intel, Compaq, and Microsoft Corporations.

Finally, Lee et al.'s paper—*Efficient, Protected Message Interface in the MIT M-Machine*—shows how a tightly-coupled parallel system, the M-machine, can integrate a network interface with a processor to provide extremely low-latency communication. In the MIT M-machine, the network interface sits directly next to the processor, and not on the I/O bus. This allows the M-machine processors to directly launch messages into the network from the processor registers and integrate communication events with processor scheduling mechanisms.

Two of us (Mukherjee and Hill) authored a recent IEEE Computer paper, *Making Network Interfaces Less Peripheral* (IEEE Computer, October 1998) that argues that SAN network interfaces should appear to processors more like memory than like a disk interface. This paper was submitted independently of this special issue and handled separately by the regular IEEE Computer editors.

Network interface design is a research topic of long-standing importance and thus there is a huge body of research and literature in each of these domains [1,2,3,4,5]. Comprehensive coverage is impractical in a single special issue. Nevertheless, we hope that this issue provides a glimpse of problems and challenges that lie ahead in the design of high-performance network interfaces.

Acknowledgments

We acknowledge the efforts of both authors and reviewers for submitting excellent papers and reviews. We received 28 manuscripts and obtained a total of 96 reviews. We also thank Angela Burgess, Michelle Saewert, and Helen Wood from IEEE Computer for providing helpful guidance during the entire process of creating this special issue.

References

1. *Hot Interconnects*. A Symposium on High Performance Interconnects, Stanford University, Stanford, CA.
2. *ACM SIGCOMM Conference*. Applications, Technologies, Architectures, and Protocols for Computer Communications. Sponsored by ACM.
3. *International Symposium on Computer Architecture* (ISCA). Sponsored by ACM and IEEE.

4. *Architectural Support for Programming Languages and Operating Systems (ASPLOS)*. Sponsored by ACM.
5. *International Symposium on High-Performance Computer Architecture (HPCA)*. Sponsored by IEEE Computer Society.

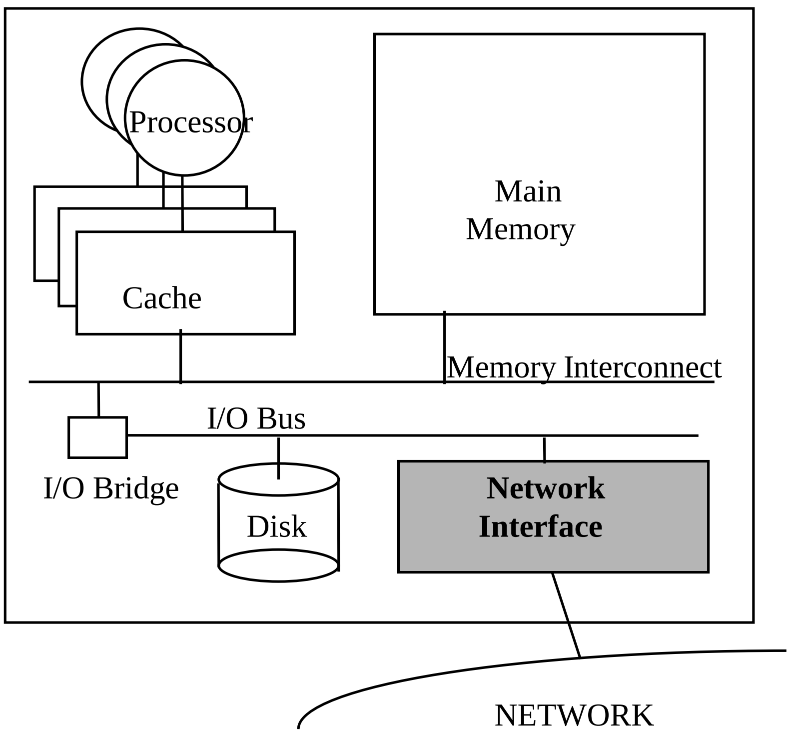

Figure 1. Architecture of a workstation node with a network interface attached to the I/O Bus.

Andrew A. Chien is the SAIC Chair Professor in the Department of Computer Science and Engineering at the University of California, San Diego and affiliated with the National Computational Science Alliance (NCSA) and the National Partnership for Advanced Computational Infrastructure (NPACI). From 1990 to 1998, Andrew was a faculty member at the University of Illinois. Andrew's research involves networks, network interfaces, and the interaction of communication and computation in high performance systems. His work also involves compilation techniques for high performance object systems. Chien received his undergraduate, master's, and doctoral degrees from the Massachusetts Institute of Technology and is a recipient of a 1994 National Science Foundation Young Investigator Award. Contact Chien at achien@cs.ucsd.edu.

Mark D. Hill is a professor and Romnes Fellow in the computer sciences department and the Electrical and Computer Engineering Department at the University of Wisconsin, Madison. He also codirects the Wisconsin Wind Tunnel parallel-computing project. His current research interests include memory systems of shared memory multiprocessors and high-performance uniprocessors. Hill recieved a BSE from the University of Michigan, Ann Arbor, and an MS and a PhD in computer engineering from the University of Califironia, Berkeley. Contact him at markhill@cs.wisc.edu.

Shubhendu S. Mukherjee is a senior hardware engineer in the Alpha Architecture team at Compaq Computer Corp. His research interests include network interfaces for system area networks, coherence protocols for shared-memory multiprocessors, and microarchitectures for high-performance uniprocessors and multiprocessors. Mukherjee received a BTech from the Indian Institute of Technology, Kanpur, and an MS and a PhD from the University of Wisconsin, Madison. Contact Mukherjee at shubu@muhthr.hlo.dec.com.