White Paper
Intel® Ethernet Controllers

Improving Network Performance in Multi-Core Systems

Responding to Next-Generation Datacenter Needs

Today's resource-intensive applications running on Intel® multi-core processor-based platforms are driving the need not only for greater networking bandwidth, but also for more efficient processing of network data. The lower power usage and greater processing capabilities of these systems has IT departments relying on them to handle more data than ever before. And with the growth of server virtualization and consolidation, high-volume data transactions, real-time technologies such as VoIP and video on demand, and the increase to 10GbE bandwidth, the processing demands on the CPU will continue to grow.

Intel's next generation of Gigabit Ethernet and 10 Gigabit Ethernet controllers support a number of features that are optimized for new Intel multi-core processor-based platforms and satisfy the networking requirements of the next-generation datacenter. Technologies such as Receive-Side Scaling (RSS), Extended Message Signaled Interrupts (MSI-X), multiple queues, and Virtual Machine Device Queues (VMDq) distribute network processing across multiple processor cores, thereby lowering CPU utilization and increasing application responsiveness. These technologies also provide significant performance improvements in virtualized server environments. This improved performance includes not only increased bandwidth, but also lower CPU utilization levels and decreased latency.

In previous-generation platforms, application data requests were associated with a single processor and handled sequentially, raising CPU utilization levels and lowering system performance under heavy network loads. Figure 1 shows a typical network data flow scenario in a previous-generation platform.

Table of Contents

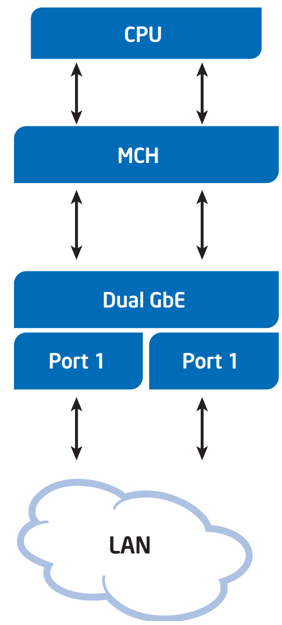

Figure 1. Network data flow in a previous-generation platform

Performance impact cannot be improved by simply increasing the bandwidth on the system. In fact, adding more network controllers only escalates the problem: larger numbers of packets cause greater processing congestion for the CPU, and latencies climb proportionately.

Solving the Problem

To address these performance issues, the Intel® 82575 Gigabit Ethernet Controller and the Intel® 82598 10 Gigabit Ethernet Controller rely on several key capabilities to distribute packet processing across the cores. These technologies work in concert to create independent packet queues, direct network packets to the correct queue, map the queue to a processor core or virtual machine, and facilitate the interaction between system, queues, and cores.

The key technologies supported by these controllers include:

- Multiple descriptor queues
- Receive-Side Scaling (RSS)
- Virtual Machine Device Queues (VMDq)
- Extended Message-Signaled Interrupts (MSI-X)

These technologies are discussed in detail in the following sections.

Multiple Descriptor Queues

Multiple transmit and receive queues in the controllers allow network traffic streams to be distributed into queues. These queues can be associated with specific processor cores, allowing distribution of the workload and preventing data traffic processing from overwhelming a single core. The packet queues can be accessed by driver threads running on different processor cores, such that multiple cores can process network packets in parallel.

Packets can be directed to individual queues in two ways:

- RSS, which has a table that maps queues to processor cores
- VMDq, which filters data into queues based on MAC address or VLAN tags

The Intel 82575 Gigabit Ethernet Controller supports four transmit and four receive queues per port. The Intel 82598 10 Gigabit Ethernet Controller provides 32 transmit queues and 64 receive queues per port, which can be mapped to a maximum of 16 processor cores. This enables considerable load balancing on servers with several multi-core processors.

Receive-Side Scaling (RSS)

To determine which receive queue to use for incoming packets, controllers residing on systems using Windows* Server 2003 or Windows Vista* can use RSS. (On Linux* systems, this technology is known as Scalable I/O). RSS directs packets to different queues without the need for reordering. Incoming packets are first segregated into flows. The specific flow for a given packet is determined by the calculation of a hash value derived from fields in the packet header. The resulting hash value serves as a look-up mechanism in a table that indicates to which flow or queue the packet should be directed. The hash values are also used to select a specific processor to handle the packet flow, ensuring that the packets are handled in order.

RSS is intelligent in its distribution of packet processing and is also programmable. Hence, by its judicious use, controllers with multiple queues can efficiently direct multiple TCP/IP streams to different processor cores for handling. The Intel 82575 Gigabit Ethernet Controller and the Intel 82598 10 Gigabit Ethernet Controller both support RSS.

Virtual Machine Device Queues (VMDq)

The ability to direct streams to different cores is also an important element in supporting virtualization. This design allows virtual machines hosted by hypervisors that emulate network controllers to rely on a dedicated network stream that is handled by a single core. Hence, when multiple virtual machines (VMs) are in use, they can share the controller ports while enjoying their own privately processed packet stream, a solution that greatly improves virtualized performance.

Intel's VMDq technology provides multiple hardware queues and offload features that can be used to reduce the software overhead associated with sharing a single networking controller between multiple virtual machines. Prior to the advent of this technology, a network switch emulated in software by the virtualization platform sorted and routed the packets individually to the running VMs. This process introduced significant delays in the network packet processing. With VMDq, individual hardware queues are associated with the simulated network interfaces of the running VMs, so the controller itself performs the routing of received packets, thereby substantially lowering the overhead. This technology is also used on outbound VM packets to provide transmit fairness and to avoid a single VM blocking access to the controller.

Extended Message-Signaled Interrupts (MSI-X)

The ability to communicate efficiently between queues and particular processor cores is handled by MSI-X. MSI-X is the next generation of MSI, which passes interrupts to a single processor core. Conversely, MSI-X provides multiple interrupt vectors, which allow multiple interrupts to be handled simultaneously and load-balanced across multiple cores. This improvement helps improve CPU utilization and lower latency.

The Intel 82575 Gigabit Ethernet Controller and the Intel 82598 10 Gigabit Ethernet Controller give each queue its own set of MSI-X controllable interrupt vectors, which permits efficient packet management and fine tuning of the processor load. With an interrupt vector for each queue, the controller can handle multiple interrupts simultaneously, preventing the bottlenecks associated with funneling all interrupts through a single vector.

Figure 2 shows how these features work together to distribute Ethernet traffic across CPU cores in a multi-core system.

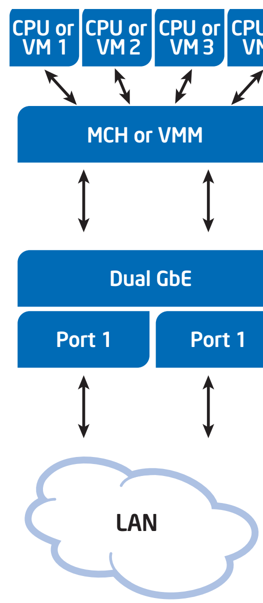

Figure 2. Multiple queues, RSS, MSI-X, and VMDq working together in a multi-core system

Intel® Network Controllers

In addition to the technologies discussed in the previous section, the Intel 82575 Ethernet Controller and the Intel 82598 10 Gigabit Ethernet Controller have unique performance-enhancing features. For example, both controllers use Intel® I/O Acceleration Technology (Intel® I/OAT) to optimize bandwidth by redirecting header processing and improving memory access to packet components. As a result, these controllers are among the fastest available today, especially when run on multi-core processor-based platforms.

Intel® 82575 Gigabit Ethernet Controller

This controller is a PCI Express*-based, dual-port, Gigabit Ethernet controller with four transmit and four receive queues per port. It supports multiple queues, RSS, VMDq, and MSI-X. In addition, it can offload IP processing and checksumming, and it can perform TCP transmission segmentation.

For more information, see www.intel.com/design/network/products/lan/controllers/82575EB.htm

Intel® 82598 10 Gigabit Ethernet Controller

The Intel 82598 10 Gigabit Ethernet Controller is a PCI Express-based, dual-port, 10-gigabit Ethernet controller, designed for very high-bandwidth needs. It has extensive support for multiple processor cores and is intended for use on large systems, such as enterprise servers, processing appliances, and in embedded applications. It provides 32 transmit queues and 64 receive queues per port. For virtualization, it offers 16 virtual machine device queues per port. To provide maximum throughput for these queues, it has a 512 KB receive buffer (which can be subdivided into eight individual packet buffers) and a 320 KB transmit buffer (also divisible into eight packet buffers). In addition, the controller offloads tasks from the host such as TCP/UDP/IP checksum calculations and TCP segmentation.

For more information, see www.intel.com/design/network/products/lan/controllers/82598.htm

Next Steps

Intel's market leadership in processors and research leadership in high-bandwidth network controllers give it the ability to create networking components that optimize the delivery of network packets to applications. In a world that increasingly relies on the network for real-time or near real-time data delivery, Intel controllers are strategic components needed for high-performance platform designs.

For more information on Intel networking products, go to www.intel.com/network or contact your Intel representative.

Printed in USA 0907/BY/OCG/PP/500 Please Recycle 318483-001US