

TOE: TCP/IP Offload Engine relieves CPU burden

By Alan Baldus



Today's CPUs are overwhelmed by Ethernet traffic, with an increasing amount of CPU cycles being consumed by TCP/IP packet processing. An additional CPU can be added to share the load when this imbalance causes application performance to suffer, but this is an expensive solution with several other drawbacks.

In this article, Alan discusses how the addition of a Network Accelerator Card (NAC) with an integrated TCP/IP Offload Engine (TOE) can relieve the burden on the CPU.

TCP/IP processing

The TCP/IP protocol stack is the common thread that links today's LANs (Local Area Networks), WANs (Wide Area Networks), and SANs (Storage Area Networks). The protocol stack uses a substantial amount of resources for protocol processing, and it has become the major bottleneck in high-speed networks. As network speeds continue to increase, the situation is only growing worse.

PCI servers today commonly integrate at least one Network Interface Card (NIC), the classic workhorse that typically provides a 1 Gb Ethernet port. A NIC only offloads a small part of the total network processing load from a CPU, and does not offload any of the TCP/IP packet processing.

Relying on the CPU to perform TCP/IP protocol processing is not necessarily the problem as long as enough CPU cycles remain available to service system applications. As a rule of thumb, it takes virtually all the horsepower of a 1 GHz microprocessor to service a single full-duplex 1 Gb Ethernet network connection. Networks are now evolving to 10 Gb Ethernet, and many processors will be completely saturated servicing a network connection. There is clearly a need for some hardware assist, especially for applications such as multimedia content delivery where revenue is directly proportional to performance.

Enter the TOE NAC, a special-purpose NAC that is specifically designed to offload the burden of TCP/IP processing

from the CPU so more CPU cycles are available to applications. There are substantial differences in current TOE NACs in architecture, performance, and CPU offload capability. Unless a NAC reduces the CPU utilization required for network processing by at least 50 percent, it is not much better than a NIC.

TOE market segments

There are four major market segments for TOE today:

- HPC (High-Performance Computing) and supercomputing platforms
- Multimedia content delivery systems (packet delivery)
- Next-generation IP storage including SAN, NAS (Network Attached Storage), and iSCSI (Internet SCSI)
- Transaction processing systems such as e-commerce servers

These segments have some common characteristics, but they differ in the relative importance of such issues as Ethernet transaction latency and the number of manageable simultaneous connection sessions.

RISC-based TOE NACs

One class of TOE NACs specifically serves the iSCSI IP storage market, but these cards typically cannot meet the needs of the other market segments. It might seem that an iSCSI NAC would suit the needs of a multimedia content delivery system since both these segments tend to deal with very large packets. However, multimedia systems require a fairly high number of sessions to be conducted simultaneously, while iSCSI requires a fairly low number of concurrent sessions.

Moreover, the iSCSI-oriented TOE NACs are usually based on general-purpose RISC microprocessors, specifically MIPS-architecture devices. While these offer a low cost of entry, RISC microprocessors are relatively limited in horsepower, and are relatively demanding in terms of board real estate and power consumption.

In addition, RISC microprocessors require a large number of instructions to conduct

even a single session. Designs based on these general-purpose processors become CPU bound when called on to manage more than a few concurrent sessions, and therefore network performance rapidly degrades. For applications that require large numbers of simultaneous sessions (such as multimedia applications), such bottlenecks can have a very dramatic and negative effect on the revenue stream.

As shown in Figure 1, NICs leave all of the TCP/IP processing to the CPU, and some NACs (such as RISC-based TOE NACs) only offload a small portion of the TCP/IP processing load from the CPU. Installing such a NAC in a high network traffic application will lead to results not much better than those provided by a conventional NIC.

ASIC-based TOE NACs

In contrast, another class of TOE NACs is based on state machine ASICs, and they provide more bandwidth and flexibility in order to meet the needs of all market segments. Unlike generic RISC processors, state machine ASICs are custom designed to process TCP/IP packets quickly and without error.

TOE NAC class performance differs markedly in the area of Ethernet packet processing and reassembly. Because Ethernet packets have a distinct size limit, larger files (such as media content files) must be broken up or *segmented* into multiple packets. With Ethernet, packets often arrive out of order, as when a packet is dropped during transmission and must be retransmitted (Figure 2).

When packets arrive out of order, less capable RISC-based TOE NACs must wait until all packets arrive before they begin reassembling the segments. Since the more capable state machine ASIC-based TOE NACs are based on an in-line processing architecture, packets are processed upon arrival and it does not matter if they are out of order.

TOE NAC test platforms

To quantify the benefits of TOE NACs, SBE recently conducted a series of sys-

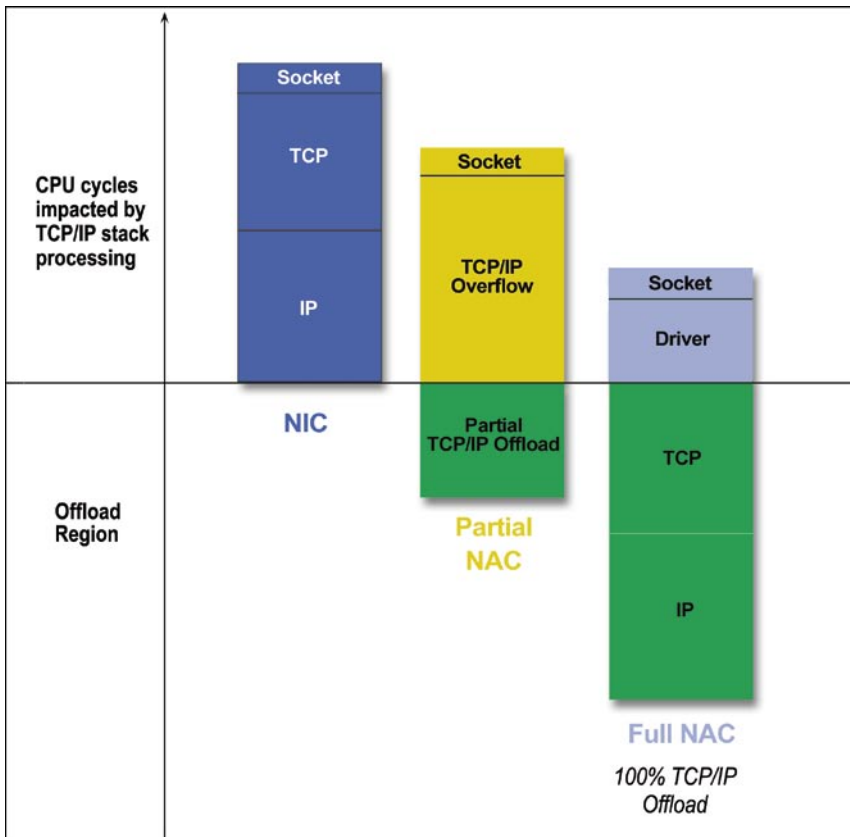


Figure 1

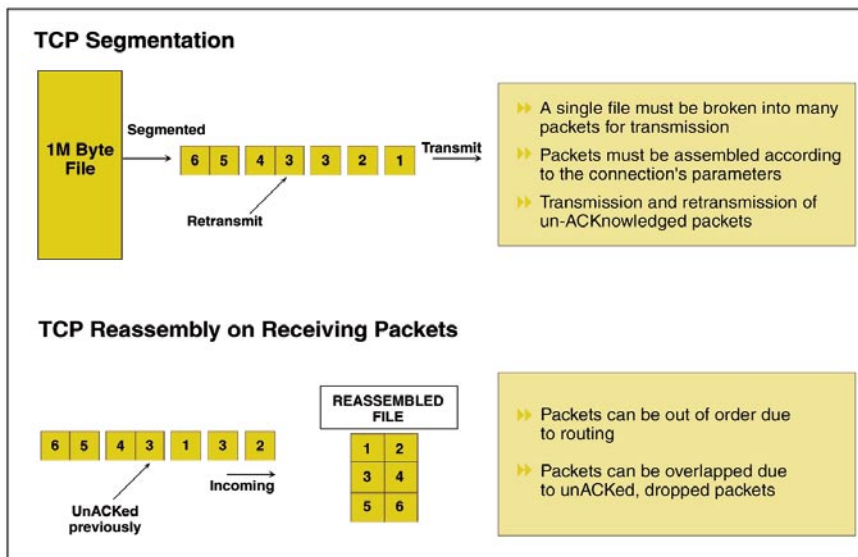


Figure 2

tem-level tests on two different PCI server configurations using the Netperf benchmarking tool (netperf.org) in a high session count environment. Netperf effectively measures unidirectional throughput and end-to-end latency for many types of networks.

The following platforms were used:

- Platform 1 – dual-processor 3.06 GHz Intel Xeon CPUs
- Platform 2 – dual-processor 1.4 GHz AMD Opteron CPUs

Both platforms included the following:

- Linux Red Hat 2.4.18
- 2 GB of system memory
- 64-bit 66 MHz PCI system bus

TOE NAC test results

Each platform was tested with and without an SBE toePCI-2Gx TOE NAC. The NAC is a half-size, 64-bit 66 MHz PCI board that includes two 1 Gb Ethernet ports. This NAC is capable of offloading 100 percent of the TCP/IP processing from a CPU. It is based on a state

machine ASIC and features dual on-board PCI buses. The TOE NAC test results are shown in Figure 3.

Without the TOE NAC (NIC only), CPU utilization for the benchmark varied from 63.96 percent for the Xeon platform, to 66.7 percent for the Opteron platform.

With the SBE toePCI-2Gx TOE NAC, the CPU utilization benchmark dropped to 14.8 percent for the Xeon platform, and 18.0 percent for the Opteron platform.

For the Xeon platform, the TOE NAC board therefore decreased CPU utilization from 63.96 to 14.8 percent, an offload factor of 77 percent.

For the Opteron platform, the TOE NAC board decreased CPU utilization from 66.7 to 18.0 percent, an offload factor of 73 percent.

It has thus been demonstrated that the inclusion of the SBE toePCI-2Gx TOE NAC freed up a substantial amount of CPU bandwidth for application processing.

Benefits example

The benefit TOE brings to systems is readily apparent in multimedia content delivery systems and other applications where performance directly impacts revenue flow.

For example, a garden variety server for the multimedia content delivery market includes a dual-processor Xeon, 4 GB of system memory, SCSI disk drives, and 1Gb Ethernet.

Operating at 50 percent of the CPU capacity with a heavy application load, this server will provide a throughput of about 30,000 Packets Per Second (PPS) with varying size packets. Full-out operation with no reserve comes in at about 60,000 PPS.

With a TOE NAC, at least 50 percent of the TCP/IP processing will be offloaded from the CPU. This will cause the packet delivery performance to increase by 50 percent from 30,000 to 45,000 PPS when using 50 percent of the CPU capacity. Correspondingly, the packet delivery performance will increase from 60,000 to 90,000 PPS at full CPU capacity.

The server in this example is currently priced at about \$4,000, so the network bandwidth cost is about 14 cents per packet at 30,000 PPS. With a TOE NAC in place, the network bandwidth cost drops to 7 cents per packet (without accounting for the cost of the TOE NAC). Meeting

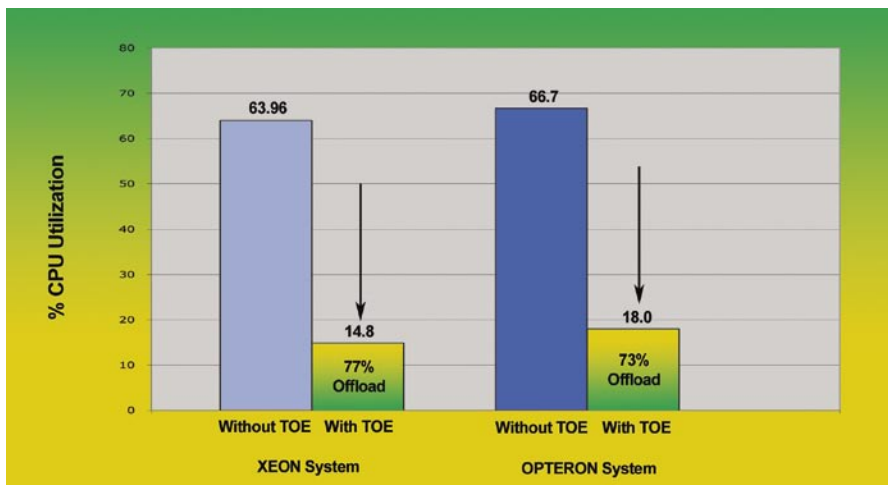


Figure 3

even more demanding cost per packet requirements opens new opportunities for TOE NACs.

Summary

When packet traffic threatens to overwhelm a CPU resulting in inadequate bandwidth for applications, the installation of an additional CPU board in a server or an additional server in a rack can solve the problem. This, however, is not an effective solution in terms of cost, power consumption, and real estate. It is also an unattractive approach for cost-sensitive installations where the leverage

of the existing hardware investment is critical.

In contrast, the addition of a specialized coprocessor in the form of a TOE NAC provides an immediate and apparent Return On Investment (ROI); and it also has a great real estate advantage over adding an auxiliary server for installations with space constraints. By offloading TCP/IP protocol processing, the TOE NAC allows users to leverage their existing hardware investment, while freeing up the CPU to process real applications.

Alan Baldus is the Director of the Field Application Engineering (FAE) team at SBE. Alan has been working in the embedded networking industry for over ten years. He has been active in the PICMG organization and chaired the PICMG 2.15 PTMC specification. Alan holds a BS in Electrical Engineering from San Diego State University.

For more information, contact Alan at:

SBE, Inc.

2305 Camino Ramon
Suite 200
San Ramon, CA 94583
Tel: 925-355-2000
Fax: 925-355-2020
E-mail: alanb@sbei.com
Website: www.sbei.com