

Microsoft Chimney: The Answer to TOE Explosion?

Saqib Jang 08/19/03

Summary

While TOE technology and market have continued to gain mindshare in the trade media since 2000, TOE has been less-than-successful from a deployment standpoint. The *all-inclusive* approach for architecting TCP offload as a parallel protocol stack has been driven by the lack of a standard operating system interface for TOE. This approach has had deployment, security, and time-to-market implications. The only TOE vendor gaining market traction has done so with a focus on purpose-built server applications, such as file service appliances, backup media servers, and medical imaging systems, where the risks of following a proprietary offload approach can be managed.

Microsoft's recently announced Chimney offload architecture (GA mid-2005) represents a significant departure from the prevalent approach of architecting TCP offload. Chimney has a number of key benefits versus the all-inclusive TCP/IP offload approach, include reduced development and deployment complexity, and focuses on minimizing the security implications of TCP/IP offload. Most importantly, Chimney will drive the availability of TOE-enabled GbE ports as a standard motherboard component in the 24-36 month timeframe. Vendors offering stand-alone TOE-based products, such as iSCSI HBAs and TOE NICs (T-NICs) will likely focus on the niche market of providing value-added TOE capabilities such as IPsec and RDMA acceleration for servers, as well as targeting the iSCSI target market with higher-end, premium-priced products.

Chimney will also have a significant impact on how TCP offload is architected in ancillary markets, such as the Linux server market and iSCSI target markets. A Chimney-compatible Linux architecture for TCP offload will emerge in 24-36 months to leverage the volume efficiencies of Chimney-compatible GbE TOE ports on server motherboards. In addition, the range of startup and established vendors offering TOE and iSCSI acceleration will shift their focus to providing Chimney-compatible ASICs and NICs for iSCSI switches and arrays with advanced capabilities such as 10 GbE support, iSCSI, IPsec, and RDMA.

TOE Technology and Market Update

While TOE¹ technology and market have continued to gain mindshare in the trade press since 2000, TOE has been less-than-successful from a deployment standpoint. There is only one vendor shipping TOE chips and NICs in volume, while a range a few other vendors have made "soft" product announcements lacking OEM and end-customer testimonials and case studies, real-world benchmarks, and other value proposition specifics.

Most TOE vendors have chosen to follow an *all-inclusive* approach for TCP/IP offload, including offload of connection-setup, data path offload, and support for ancillary protocols such as DHCP, ARP, ICMP, and IGMP. Vendors have believed that such an approach was required to meet the goal of gigabit-level wire-speed performance, especially for latency sensitive applications such as IP block storage, as well as to address the lack of standard interface enabling integration of TOE capabilities with server operating system TCP/IP protocol stacks.

In general, vendors are shipping TCP/IP offload as a parallel protocol stack. On Windows, these folks are called "TDI Hookers". This is because they "hook" into the Windows Transport Driver Interface (TDI) API that is at the top of the transport stack, enabling their software to intercept all networking communication

¹ For the purposes of this discussion, TOE refers to TCP offload capabilities embedded in protocol-specific products (e.g. iSCSI HBAs) as well as general-purpose TCP acceleration products (e.g. TOE NICs).

with the application. Alacritech's approach falls into this category. Another approach is to masquerade as a "storage device" - make the storage stack thinking that the its a Host Bus Adapter (HBA) talking SCSI protocols, and the fact that underneath the covers the product is offloading iSCSI and TCP is not apparent to the software stack. Either approach requires the implementers to duplicate all of the functionality mentioned previously, and has a number of important weaknesses pertaining to development and deployment complexity, and network security.

First, it leads to administrative complexity of managing two different TCP/IP implementations within the same system, each of which has a separate repository for connection-state and protocol-specific information. In addition, offloading of connection set-up may have security implications. By having centralized connection management in operating system software that's been hardened and tuned over a number of years, there is a single, mature location to protect against denial of service attacks and other attacks, along with typically much more CPU resource to perform this protection.

TOE vendors have typically focused on an ASIC-based approach to implement the full TCP/IP protocol stack. While the performance gains of such ASICs can be orders of magnitude greater than a general-purpose processor-based approach, it presents a corresponding increase in hardware development complexity and cost, which explains why most TOE vendors have struggled in shipping their products and why the price points of such products (\$500+ per GbE TOE NIC vs. \$50 per GbE port) are in the way of volume deployment.

It is interesting to note that Alacritech, the only TOE vendor gaining significant traction especially in the dedicated server applications (e.g. file serving appliances, backup, video editing, and medical imaging), is not following the all-inclusive protocol stack approach. Its approach consists of only offloading the TCP/IP data path to hardware, while connection management and state-based operations, such as retransmissions and fragmentation, are implemented in server-resident software drivers. However, the lack of a standard OS interface for TOE integration continues to be an issue for Alacritech as its T-NIC driver for Windows servers breaks a number of applications² (e.g. packet filtering firewalls, network address translation, network load balancing etc.), including management applications. While this may not be an issue for function-specific server applications, it presents significant complexity for general-purpose deployment.

Microsoft Chimney Overview

Microsoft's announced its Chimney Offload Architecture for Windows in May 2003 at the May '03 WinHEC. The goal behind Chimney is to provide a standard interface for integration of TOE products with the Windows OS stack³ for simplified development and deployment of Windows-based TOE products.

The Chimney architecture represents a significant departure from the prevalent all-inclusive approach of TCP offload. Specifically, Chimney includes offload of TCP/IP connection data path to the offload hardware (or target), such as an iSCSI HBA or a TOE NIC, while set-up and teardown of accelerated TCP/IP connections and support of ancillary protocols such as DHCP, RIP, IGMP, and ARP will be implemented within the Windows TCP/IP stack. Further, Chimney allows the offload of retransmissions, but not IP fragmentation. There are a number of operating systems that generate IP fragments, and there are several security attacks with them. Thus, having the host stack do the reassembly and forward the re-assembled TCP segment to the NIC ensures protection against vulnerabilities.

There a number of important benefits of the Chimney approach of TCP/IP offload vis-à-vis the all-inclusive-approach of TCP/IP offload. First, the use of the Windows TCP/IP stack for supporting connection set-up/tear-down and support of ancillary protocols enables a simplified deployment model and

²While Alacritech includes an algorithm to revert to the MS TCP/IP for unsupported features, such an approach has limitations. First, a number of protocols (e.g. NFS, SMB, iSCSI) use unpredictable port numbers on the client. Second, the 'hit and miss' nature of applying this approach make it a challenge for enterprise-level deployment.

³ A presentation on Chimney can be found in the Windows Architecture for Scaleable Networking session at the WinHEC 2003 site @ <http://www.microsoft.com/whdc/winhec/pres03.mspix>,

benefits of TCP/IP acceleration to be made easily available to all Windows-based network applications, including iSCSI. For example, the Windows iSCSI initiator service, which uses the Windows TCP/IP stack, will automatically utilize the acceleration benefits of a Chimney-enabled TOE NIC. Second, support of connection set-up by the Windows protocol stack addresses the security vulnerabilities posed by offload hardware handling connection set-up. Third, the Chimney offload model significantly lowers the complexity of developing TCP offloading ASICs and NICs for the Windows server market.

Microsoft's publicly-announced plans are for Chimney to be part of the next major release of Windows server code-named Longhorn. Latest estimates are for two Longhorn beta releases during 2004 and Longhorn general availability to be in mid-2005.

TOE Market Implications

Chimney will have a number of important implications for the overall growth of the TOE market as well as for strategies of TOE vendors. First, I believe that Chimney will drive GbE TOE to be a high-volume standard server capability over the 24-36 month timeframe. It is noteworthy that Broadcom was one of two vendors (the other being Adaptec) to announce support for Chimney at Microsoft's announcement of the new architecture at the spring '03 WinHEC. Chimney is ideally suited to allowing TOE to be a standard function for server GbE ports benefiting server motherboard component vendors such as Intel and Broadcom. The reduction in TOE complexity driven by Chimney will make it easier for such vendors to include TOE capabilities "for free" in their next-generation GbE controller components. In addition, I believe that the Linux server market will evolve to support a similar type of offload model as Chimney so that Linux can leverage Chimney-enabled server motherboard-based GbE TOE capabilities.

With the picture looking very clear for GbE TOE to become a standard server motherboard capability, what does portend for the range of vendors planning to offer products TCP offload products for standard Windows and Linux servers, including iSCSI HBAs and TOE NICs? It is an open issue whether such the performance value proposition of such products is sufficiently better vis-à-vis Chimney-enabled server motherboard-based GbE TOE capabilities. A more likely scenario is for such vendors to target a limited, niche market for premium-priced iSCSI HBAs and TOE NICs providing value-added capabilities such as dual-porting, IPsec, and TCP/RDMA (the Chimney architecture includes interfaces for offloading of IPsec and TCP/RDMA capabilities).

A more likely scenario is for the availability of motherboard-based TOE capabilities to drive vendors developing all-inclusive iSCSI offload products to focus on the market for iSCSI targets, include iSCSI switches and arrays. The performance requirements of iSCSI and TCP termination of iSCSI targets will be significantly higher than for iSCSI-enabled servers. In addition, priority of requirements such as security and 10GbE support will be higher for this market compared to the standard server market. While iSCSI targets will typically use embedded versions of Linux or proprietary embedded operating systems, a Chimney-compatible model of offload architecture will typically be used to enable TCP offload.