Practical Innovation

A White Paper from Arroyo Video Solutions

The 10GigE Migration

Designing 10GigE-based Transport Networks for IP Video and VOD

By Paul W. Sherer, President & CEO, Arroyo Video Solutions

June 2005

Summary

Only a few years ago, 10GigE seemed at best a vision for the future. But today it is widely available in NIC interfaces for servers, within switches, and supported by wideband fiber transport infrastructures for both local and remote intersystem communications applications.

Within the same timeframe, video services have migrated rapidly away from the original point-to-point architecture of ASI transport and TDM-over-fiber, to GigE transport. IP video and video-on-demand (VOD) over GigE infrastructures is now a widely deployed service.

The availability of such high speed infrastructure presents novel opportunities for the design of systems carrying IP video and VOD. With such flexibility, subscriber "churn" will be reduced and MSOs will be able to architect solutions vastly richer and more functional than ever seen in the past.

This paper retraces the evolution toward 10 GigE switches, interfaces, and transport. 10GigE enables a new array of architectural methods for designing and deploying IP TV and VOD service offerings — and is readily migrated to other forms of real-time and near real-time applications.

The Evolution from ASI to 10GigE

When Ethernet got its start in the early 1980's, the market was very confused about transport standards for local area networks. IBM Token Ring, Datapoint ARCNET, and a potpourri of other standards competed for dominance as network interface standards. Ethernet itself started off with slow data rates — 2Mbps, 5Mbps, and 10Mbps. All media were shared access media. And the debate was dominated by arguments over whose media access control method gave the best control over bandwidth and priority. In the late 80's, 3Com and other vendors completely redefined the landscape with the invention of the "collapsed backbone" concept.

A collapsed backbone is defined as the use of the backplane of a switch to network together point-to-point forms of Ethernet. The capacity of the backplane is orders of magnitude higher than the bandwidth of the physical interfaces. And solutions began to appear that were non-blocking in the backplane.

Meanwhile work appeared at the transport layer — metro Ethernet over Fiber, ring networks, and point-to-point networks that essentially "uncollapsed" the backbone. Standards evolved adding services such as virtual LANs, priority, multicast features, and link aggregation (n*ETH). Costs were reduced and feature-rich Ethernet became the mass market media interface, with other competing standards rapidly dropping out of the game.

Today Ethernet is the dominant form of enterprise, metro, and (increasingly) backbone technology. While the traffic is IP, the interfaces are increasingly standardized GigE and n*GigE. Ethernet is used in over 85%

of all network connection interfaces, and over 300 million hub and switch ports have been installed.

At the same time, the video industry migrated from analog to digital and defined its own industry-specific set of interfaces and standards — in particular ASI — for the transport of real time video. Figure 1 outlines the structure of systems using ASI.

The industry evolved ASI. Transport products were developed that carried ASI in a branching tree topology from the headend to the edge. Variants were evolved that carried ASI over metropolitan rings as an out-of-band service, and encoders and decoders were developed with ASI interface cards.

On the surface, things appeared to work out well, at least for broadcast-only applications. But as video evolved to a more switched model (multicast groups), and towards an on-demand model (unicast), the ASI networking toolkit did not scale accordingly.

A comparison of what can be done with Ethernet (and IP) systems vs. what can be done with ASI systems is particularly interesting:

- Ethernet scales to n*GigE data rates; ASI scales to sub-gigabit data rates.
- Ethernet has standardized switching solutions (which are also compatible with IP routers); ASI has limited point-to-point solutions, and no switching capabilities.
- Ethernet supports unicast, multicast, and broadcast addressing; ASI is unicast only.
- Ethernet supports very flexible switching topologies, such as tree and branch, ring, and arbitrary topologies; ASI supports simple point-to-point topologies.
- Ethernet supports VLANs — isolated broadcast domains; ASI does not.
- Ethernet supports priority tagging with up to 8 priority levels; ASI does not.
- Ethernet supports jumbo frames up to 9Kbytes; ASI does not.
- Ethernet economics and performance are driven by global mass market drivers; ASI is a tiny and expensive market, of very limited performance.
- Ethernet innovation rate is high and maps to server, switch, and transport innovation; ASI's innovation is limited and is primarily focused on how to "bridge" from legacy ASI to next generation GigE Ethernet.

10GigE products, both NICs and switches, are being driven rapidly by the enterprise server market. Volumes are tripling year to year. Pricing is currently dropping to under $5K per switch port and under $1K per NIC card. As such, 10GigE is no longer a technology of the future, but rather a viable alternative that allows IT managers and MSO operations personnel to take advantage of the fat pipes for the consolidation of I/O ports and cabling on servers and switches.

Applications Design Using 10GigE

A design trend that started with GigE is now accelerating with 10GigE. High speed, low latency

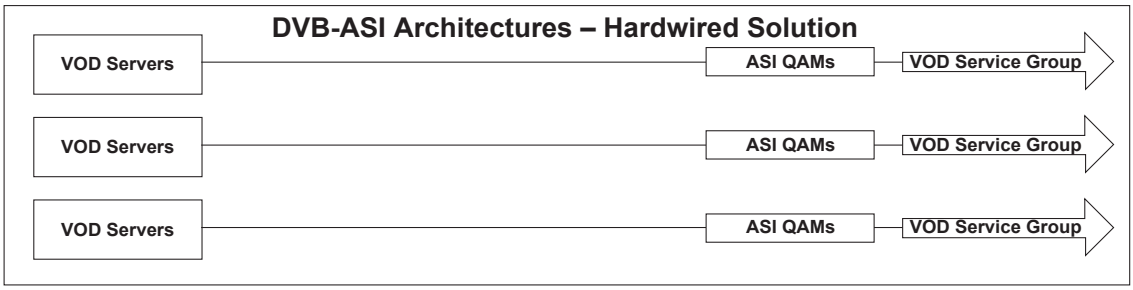

Figure 1: ASI Architectures

transport and switching structures enable the network to be utilized as the backplane of a distributed solution. This design trend started with GigE and is now acceleration with 10GigE, and is in stark contrast to the old view of systems design, in which the architecture approach was box-centric, with boxes connected together with slow point-to-point interfaces.

Some of the features of Ethernet switching (n*GigE and 10GigE) which affect network-centric solution designs include the following:

- High bandwidth can be used to assure rapid completion of high priority functions. Start-up bursts for cache-fill operations enable short latency start-up of real time video stream to "eyeballs". Also, high bandwidth can be used to rapidly enable resiliency functions such as propagation of duplicate copies of streams to other servers within the streaming server group.

- VLAN broadcast domain isolation can be used to minimize traffic flooding and also to implement clustering functions (e.g., VLAN-specific video streaming group).

- Multicast distribution can be used for keep-alive heartbeats and other resiliency functions.

- Jumbo frames can be used to minimize packet overhead and maximize throughput (e.g., performance is often limited to "packets-per-second" in servers).

- Priority tagging and switch support can be used to handle different priority network functions (such as high priority for transporting active video to a consumer, and lower priority for transporting redundant copies of the video to backup servers).

GigE and 10GigE Enables Tiered Caching

Caching is a well known method for optimizing the mix of high and low cost components in an end-to-end system. The most expensive but highest performance storage is RAM, while the least expensive storage is disk drives. High speed connectivity between resilient elements, with the option of locating those elements in different locations, enables resiliency.

In order for a caching architecture to be effective, it must contain a suite of reactive algorithms that dynamically tracks changes in content popularity. For VOD items to be tracked, they have to include the titles that are being viewed, and the most popular points within the titles that are being viewed. Popularity often cannot be predicted; a good example is the sudden death of a major entertainer, leading to a sudden uptake of libraries of content featuring that entertainer.

In addition, new forms of navigation can be enabled within caching streaming servers, such as chaptering (allowing entry into a piece of content at a set of locations) and linked virtual assets (program segments, advertisements, etc). Some chapters or virtual assets may be much more popular than others, and thus remain in cache.

Four tiers of caching are logically part of a VOD solution (Figure 2).

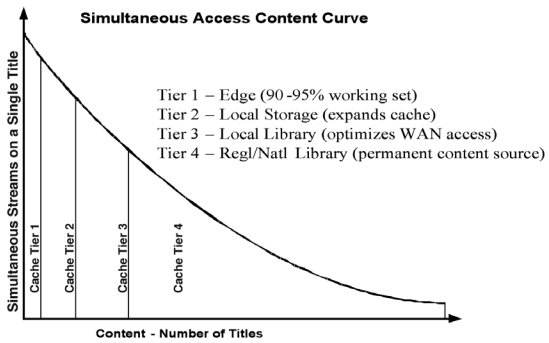

Figure 2: Caching Hierarchy

Tier 1 is closest to the edge and is where the content is played out in real-time. It also accounts for a small percentage of the hardware because of the caching efficiencies delivered by dynamic popularity algorithms.

Tier 2 is local storage and may be within a single system, or in a peer system connected by high speed local bandwidth.

Tier 3 is a local content library which enables popular content to be held locally and accessed via inexpensive networking resources.

Tier 4 is the permanent source of content and can be located anywhere, even in a national headend.

Each caching tier has resiliency within groups of systems at that same tier, and with systems at higher tiers within the architecture. The operator does not need to statically assign popularity tags to content, or manually propagate content to libraries or specific caching nodes.

10GigE (and n*GigE) reduces the distributed node-to-node transport latency sufficiently so that a variety of cache management operations can be utilized. Figure 3 shows the dramatic latency reduction enabled by 10GigE — a one-hour show can be filled in 1.35 seconds, or 2,667 independent and unique shows can be filled in one hour. These performance numbers assume full utilization of the bandwidth. Actual numbers will vary because of the multifunction use of the transport and packet overhead.

High speed transport and switching can be used as the vehicle of special operations, such as creating resiliency for the most popular titles. Since content is being filled into caches much faster than real-time, sufficient bandwidth exists to perform other management operations as lower priority operations. Functions such a "provision multiple", whereas a stream is composed of several independent streams (e.g., program plus subscriber targeted ads), can also be assembled from multiple content locations.

10GigE-based VOD Systems

Using n*GigE and 10GigE operators can build nationwide networks today. 10GigE switches and transport are used both as a clustering mechanism within hubs and as a transport mechanism between hubs. Figure 4 outlines a typical design of such a system.

The end-to-end solution consists of groups of vault servers (with thousands of hours of content stored in each vault), intermediate groups of caching servers (whenever latency reduction or bandwidth mitigation is needed), and edge groups of stream-

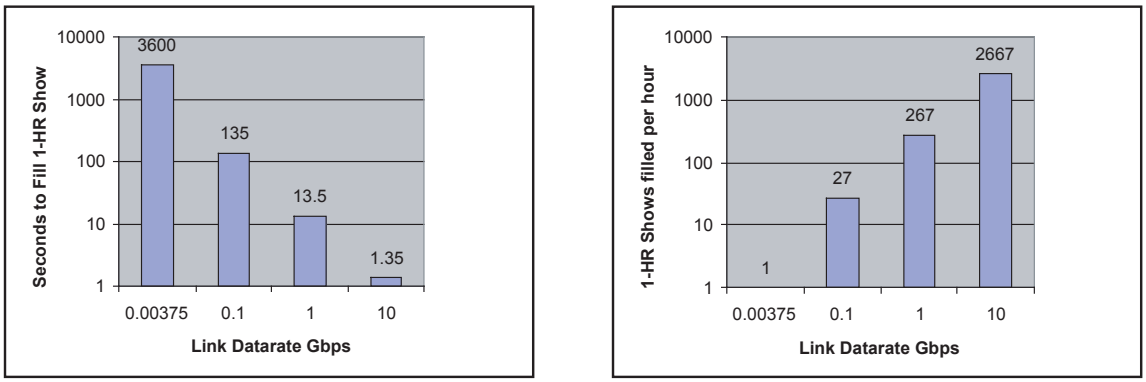

Figure 3: 10Gbps Fill Rates

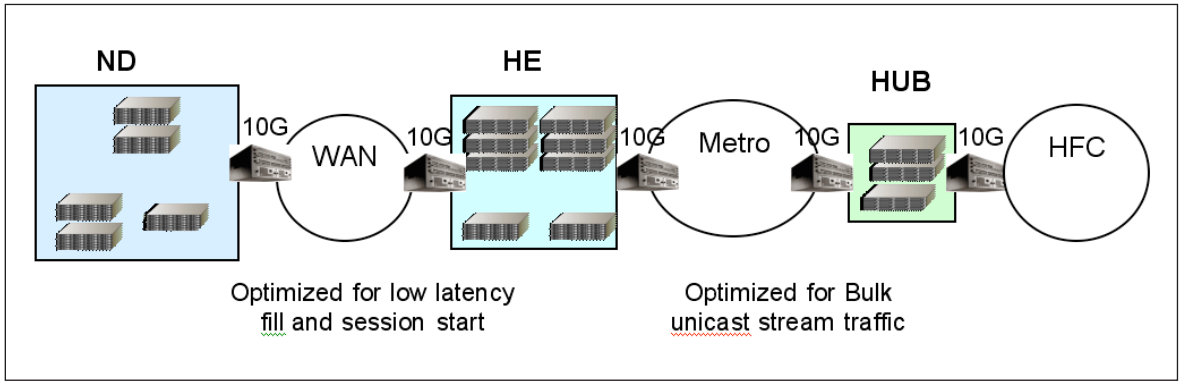

Figure 4: 10GigE-based VOD System

ing servers. The addressing and switching tools provided by the GigE switching system are utilized, such as jumbo frames, VLANs, multicast pruning of server clusters, and prioritized transport. High speed connectivity also strengthens the fault tolerance of the end-to-end system. For example, vaults and streaming caches can be replicated and rebuilt in near-real time. New systems can be added transparently, and content can be dynamically load-balanced as a background (but low latency) operation.

Finally, even the edge network interface can benefit from 10GigE transport. New breeds of edge modulators are being considered that can aggregate up to 128 6Mhz channels of downstream content. Using today's modulation (64/256 QAM), the aggregate bandwidth needed to serve such a cluster is on the order of 5Gbps. Using future modulations the spectral density can double, and thus the deployment of 10Gbps in the backbone and within streaming servers enables the flexibility to grow into future 10Gbps service capacity tiers in the edge network.

Commodity Hardware Solutions

Figure 5 outlines the tradeoff between proprietary hardware and commodity hardware. At a specific point in time, proprietary hardware can provide a higher performance solution to the market; in the early 90's, such solutions were developed for VOD. But streaming prices were orders of magnitude of what they are today, so these solutions were not ready to scale.

Today, thanks to the rapidly accelerating pace of processor performance, address space expansion, increase of on board memory, increase of I/O bus speed, and deployment of high speed NIC adapters such as 10GigE, the capabilities of standard servers has surpassed that of proprietary hardware solutions.

It is now feasible to use off-the-shelf servers to process and stream 10GigE's worth of streaming content. In addition, 10GigE can be used to handle key system functions such as content propagation and resiliency within a distributed system.

Because of the rapid product development cycles within the commodity server market, it is now also possible to develop a deployment strategy based on commodity hardware, and use software to enhance and differentiate service. Proprietary hardware solutions never leverage the volume economics of the commodity market, and lag behind commodity solutions in the adoption of new technology architectural innovations such as 10GigE.

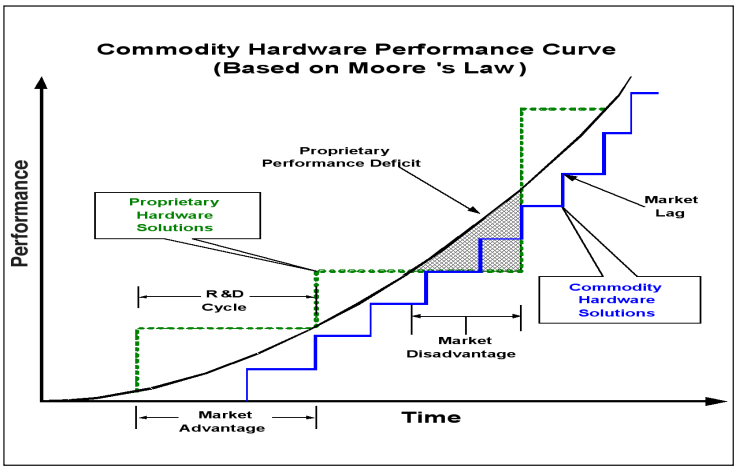

Figure 5: Commodity Hardware Performance Curve

Conclusions

10GigE and rapid advances in server hardware are enabling operators to rapidly evolve their solution architectures for deploying scaleable IP TV and VOD services. Development cycles are much faster than proprietary hardware solutions, and hardware selection is also much more flexible.

Switch, NIC and motherboard hardware prices will drop rapidly over the coming months and years. High speed, standards-based Ethernet switching architectures are already enabling network-centric, scaleable and fault tolerant solutions, giving operators the flexibility to deploy a range of architectures, from highly centralized to highly distributed, all feeding a high capacity edge network at the access edge.

Contributing companies' information

Arroyo Video Solutions

Arroyo Video Solutions is the developer of a unique video platform that enables broadband operators to deliver advanced video applications. Its first product, Arroyo OnDemand, offers a unique network-centric architecture for VOD services that uses 100% open-standard hardware platforms. Arroyo OnDemand employs Arroyo Video Accelerator™ technology, which improves performance ten-fold over other leading VOD products, and is the first solution to deliver "Always ON" nonstop VOD. Headquartered in Pleasanton, California, Arroyo is a privately held company, funded by leading Silicon Valley venture firms Matrix Partners, DCM-Doll Capital Management and Foundation Capital; and by Time Warner Investments and Comcast Interactive Capital, the investment arms of two of the nation's largest media companies. For more information, please visit www.arroyo.tv

Neterion

Founded in 2001, Neterion Inc. has locations in Cupertino, California and Ottawa, Canada. Neterion delivers 10 Gigabit Ethernet hardware & software solutions that solve customers' high-end networking problems. The Xframe® line of products is based on Neterion-developed technologies that deliver new levels of performance, availability and reliability in the datacenter. Xframe, Xframe II and Xframe E include full IPv4 and IPv6 support, and comprehensive stateless offloads that preserve the integrity of current TCP/IP implementations without "breaking the stack." Xframe drivers are available for all major Operating Systems, including Microsoft Windows, Linux, Hewlett-Packard's HP-UX, IBM's AIX, Sun's Solaris and SGI's Irix. Neterion has raised over $42M in funding with its latest C round taking place in June 2004. Formerly known as S2io, the company changed its name to Neterion in January 2005. Further information on the company can be found at www.neterion.com

Arroyo and Neterion are Technology Partners. Arroyo's Video-On-Demand solution has successfully passed a comprehensive suite of interoperability and compatibility tests with Neterion's Xframe® 10 Gigabit Ethernet adapters.

www.neterion.com

Neterion, Inc.
20230 Stevens Creek Blvd.
Suite C
Cupertino, CA 95014
Main Phone: 408.366.4600
Fax: 408.366.4650

Neterion, Corp.
349 Terry Fox Drive
Kanata, Ontario
Canada, K2K 2V6
Main Phone: 613.271.3730
Fax: 613.271.3758

Information: info@neterion.com
Sales contact: sales@neterion.com