CONSIDERATIONS IN SCALING VIDEO-ON-DEMAND SYSTEMS

Submitted to the

Society of Cable Telecommunications Engineers

Cable-Tec Expo

June 15-18, 2004

Mark Jancola

Vice President, Advanced Products & Technology

MidStream Technologies, Inc.
777 108th Ave NE
Bellevue, WA 98004
(425) 201-2022
mark@midstream.com

Legal Notice

Introduction

To compete effectively with satellite and other forms of media delivery, cable operators must continue to invest in expanding service categories and quantity of video titles available on demand. More content, and popular content, will drive take rate growth, much the same way that adding screens to a movie theater will attract more customers.

Today's 1st generation installed base of Video-On-Demand (VOD) systems must be carefully evaluated when it comes to expansion of storage and/or streaming capacity. A detailed understanding of capabilities and limitations enables sound decision-making.

This paper will equip the reader with a solid understanding of VOD system architecture alternatives and identify a complete set of factors to consider when installing a new system or growing an existing one to meet customer demand.

Definitions

Video-on-Demand (VOD) began as the ability to request and control the playback of specific movies. Since those early days, the breadth of content packaging has expanded to include Subscription VOD, Free on Demand, High Definition content on Demand, digital broadcast capture (dynamically capture content and store it in the network: NDVR), and many others. Although customer usage of each service is slightly different, the term video-on-demand (VOD) used within this paper encompasses all forms of on-demand service offerings.

Considerations in scaling VOD depend on the current and planned deployment architecture. The industry has settled on three VOD deployment architectures that will be referred to throughout this document. These deployment architectures (shown in Figure 1) are Centralized, Decentralized, and Hybrid.

Centralized VOD

A centralized system contains all streaming and storage devices at a single location central to the entire network (the Head-End). All stream requests are fulfilled from, and all title library storage is installed at, one location. Centralized systems optionally locate all distribution processing functions (including QAM modulation and RF upconversion) at the head-end or in each hub.

Centralized deployment architectures have the advantage of single-site maintenance and low storage costs (because all streaming comes from one storage repository). Content propagation issues are minimized and no requirement exists for propagation bandwidth.

On the other hand, centralized deployment architectures must grapple with the risks of having all equipment located a single point—a critical problem at the head-end could disrupt service to the entire customer base. Also, on-demand transport bandwidth requirements are the highest because streams originate from one location and must be transported to all end points in the network.

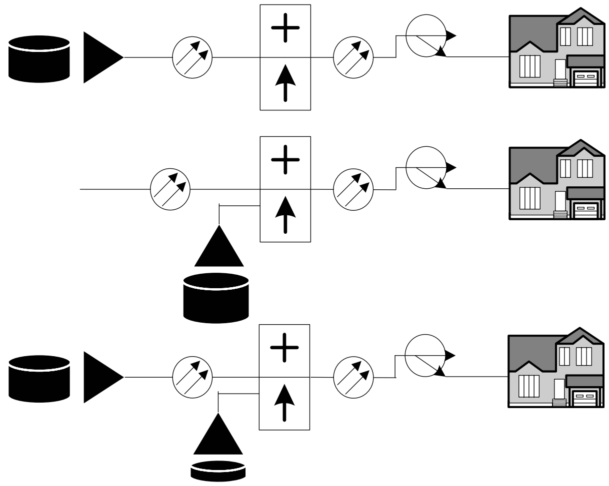

FIGURE 1. COMPARISON OF DEPLOYMENT ARCHITECTURES

Decentralized VOD

A decentralized system provides more than one location (hub sites) for streaming and storage devices. Stream requests from a particular service group are typically authenticated centrally yet fulfilled only by the devices dedicated to that service group. The entire title library is duplicated at each location.

Decentralized deployment architectures have the advantage of distributing the load across many locations such that a critical problem only disrupts service to a subset of the customer base. In addition, the on-demand stream transport network from Head-end to Hub is not necessary—it can be replaced with a relatively low bandwidth network for content propagation.

Decentralized deployment architectures have the disadvantage that VOD server storage costs can be quite high because the title library is duplicated at each location. Equipment infrastructure costs increase as decentralized deployments are less efficient due to over-provisioning (unused stream capacity cannot be redirected to other service groups). Additionally, management and troubleshooting of the network requires greater effort because more components exist in the fulfillment chain.

Hybrid VOD

A hybrid system balances the positive attributes of centralized and decentralized systems. Infrequently accessed titles, which constitute the bulk of content libraries, are served from a central location on the network. The relatively small, yet highly popular (and profitable) set of blockbusters and high demand content is served from distributed locations closer to the "edge" of the network (and closer to the customers who demand them).

Hybrid deployment architectures have the advantage of balancing the cost of storage against the cost of an on-demand stream transport network. Hybrid systems by definition are more flexible and redundant in operation.

Hybrid deployment architectures have the disadvantage that intelligent schemes for "moving" or "replicating" content from the central location to the edge (and back again) do not yet exist in robust, scalable form (although many excellent analyses have been conducted, particularly by Cox and Charter). Storage costs are higher than centralized networks as popular content is duplicated at each remote distribution hub. Furthermore, Hybrid deployments require equipment in many locations which places demands on space, power, and maintenance.

VOD Server Architectures

When we think of accommodating growth in VOD we usually think first about adding more streaming capacity. After all, revenue and return is generated on a per stream basis. Adding more streaming capacity seamlessly, and without service interruption, is crucial to maximizing the return on investment. The two components that represent the bulk of VOD system costs are Streaming and Storage.

Streaming (or output) capacity is of course tied to the number of output interfaces and the necessary processing logic to saturate them. The processing logic takes many forms, from software running on general purpose CPUs to custom hardware built precisely (optimized) for the task.

From an architectural perspective, the expansion of the library title set (storage) is even more important. Why? In just the way that the amount of water a garden hose can provide is directly tied to the water pressure, so too the amount of streams that a server can provide is tied to the performance of the storage system (bandwidth) providing the digital input bits.

Storage includes two important attributes:

1. **Storage Capacity** refers to the physical amount of content that is retained in the system. For example, a 146GB fibre-channel (FC) disk drive has an effective storage capacity of approximately 140GB or 82 hours of standard definition content (3.75 Mbps encoding).
2. **Storage Bandwidth** refers to the Input/Output (I/O) capability of a storage unit. At the hard drive level, the 146GB FC disk drive typically can sustain up to 25 MBytes per second, or enough to fulfill just over 50 standard definition streams. Storage Bandwidth at the drive level is highly dependent on the access pattern into the drive. When aggregated across many storage devices the Storage Bandwidth trends asymptotically toward the storage bandwidth of the interface link. In a 2.0 GHz FC-AL system de-rated for overhead, this is approximately 425 standard definition streams.

“Scaling VOD” refers to the scalability of both the storage and streaming/output systems. We will refer to the minimum footprint of streaming and storage to be the “**VOD building block**.” Each vendor has a unique method for defining and hooking these building blocks together. We must consider how the VOD building blocks are interconnected in order to understand the tradeoffs of near and long term scalability.

We divide the domain of VOD solution architectures into four categories: **Basic Interconnect**, **Intelligent Interconnect**, **Distributed Component**, and **Intelligent Component**.

Basic Interconnect and Intelligent Interconnect architectures utilize an interconnect scheme that is hidden within the actual VOD building blocks (streaming and storage components). The operator does not have visibility into this interconnection and therefore cannot optimize it. We call these solutions “Interdependent” because in some degree each of the building blocks is dependent upon other building blocks for normal operation.

Distributed Component and Intelligent Component architectures utilize an open system between the VOD building blocks. The operator has awareness of, and can tune if desired, the interconnecting communication logic between the building blocks. Component-based architectures, as defined herein, offer greater flexibility (the precise number of building blocks necessary are used), scalability (interconnection links can be upgraded), and increased reliability (each component is largely “independent” from others for normal operation).

VOD Systems with Interdependent Building Blocks

Basic Interconnect

In a Basic Interconnect scheme, each of the VOD building blocks talks to the other during normal operation. Content is typically striped across the building blocks so as to maximize storage bandwidth and optimize storage capacity. Figure 2 reveals the mesh communication link between each of the building blocks. Note that each block has its own output port.

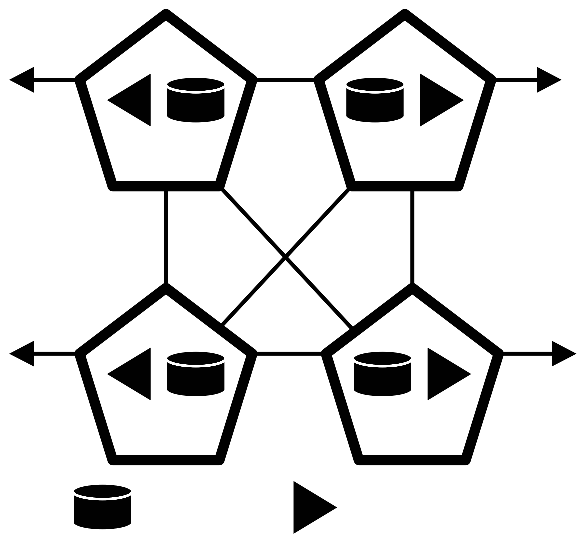

FIGURE 2. BASIC INTERCONNECT ARCHITECTURE

The interconnection link may be a proprietary or standards-based link. Often, the performance details of the interconnection link are not readily determined which means that operators will have a difficult time calculating the scalability limits.

When a stream request comes to the system, a command is sent across the interconnection link requesting each building block to fulfill a portion of the task. The results are aggregated (and collated) at the requesting block and are then transmitted.

This stream fulfillment process is similar to asking all chefs in the kitchen to prepare their portion of a meal at the same time (in parallel) and send it to the waiter as quickly as possible.

The Basic Interconnect scheme distributes the work load across numerous building blocks and thus can offer performance, redundancy, and ease of management over a specific range of streaming and storage configurations.

The primary disadvantage of the Basic Interconnect scheme comes when systems have to scale beyond the level at which the interconnect link becomes saturated.

For a simplified yet revealing example, consider a system where seven building blocks are interconnected such that the aggregated link between any two of them is dual Fast Ethernet (200 Mbps). Based on normal Ethernet limitations we can assume saturation occurs at 80% or 160 Mbps on the aggregated link. In an MPEG-2, standard definition encoding cable environment, a media packet must be transmitted every 2.8 milliseconds per stream. Each media packet is 1316 bytes constructed of seven 188 byte chunks. To assemble the packet, six 188 byte chunks must be pulled in from other building blocks, added to a seventh, and then the packet is ready. Using very general numbers, this means 1128 media bytes must be collected from other building blocks within 2.8 milliseconds across the interconnect link (160 Mbps, or 160 Kbits/millisecond). At full speed, the chunk of 1128 bytes (9024 bits) can be assembled in 0.056 milliseconds under ideal conditions. Plugging this back into the equations reveals that this system's interconnect link will saturate at *50 streams per building block for a total of 350 streams*.

To scale beyond this point, each additional building block requires the operator to add more interconnect bandwidth at each of the other building blocks or the system cannot serve more streams!

The calculation above does not consider another use of the interconnect bandwidth: content ingest into the VOD system. Active writing to the meshed storage significantly reduces the shared interconnect bandwidth further degrading stream capacity.

Therefore, existing VOD deployments that utilize a Basic Interconnect architecture have significant stream scaling problems unless the interconnect bandwidth can be dramatically increased.

The same problem exists, although in a different manner, when scaling storage requirements. Additional VOD building blocks provide greater storage, however content is increasingly fragmented causing increasing number of transactions per second across the interconnect link. As such, eventually the link will become saturated even though maximum streams remained constant and only storage increased. When an operator reaches this point, the entire system must be replicated, or replaced, just to continue scaling VOD service.

Furthermore, because the system is constructed of numerous building blocks that must communicate to function optimally, any failure point will degrade the system performance to some degree. A double failure can bring the system down entirely.

Finally, and perhaps most important, since each of the VOD building blocks contain streaming and storage functionality, operators must pay to increase both attributes even though only one may be needed. To upgrade a 2,000 stream / 1,500 hour VOD system with a Basic Interconnect architecture to 2,000 streams / 2,400 hours will require adding more stream capacity even though no additional streams are required. The cost per stream remains constant even though capital dollars are wasted on buying streams when only storage is required.

Intelligent Interconnect

The Intelligent Interconnect scheme utilizes the Basic Interconnect as its foundation yet adds interconnect logic to optimize the communication. Each of the VOD building blocks must communicate with others for optimal performance, but the intelligence allows for the condition that not every building block must be part of every transaction. Further intelligence may include a protocol such that building blocks not involved do not handle the request—it is simply passed through to the next node.

The diagram in Figure 3 reveals a more careful interconnect scheme: each building block is connected only to a few neighbors rather than in a full mesh. Such a scheme reduces cost (in terms of physical interfaces) and requires an intelligent low latency routing protocol for commands and data to traverse the network.

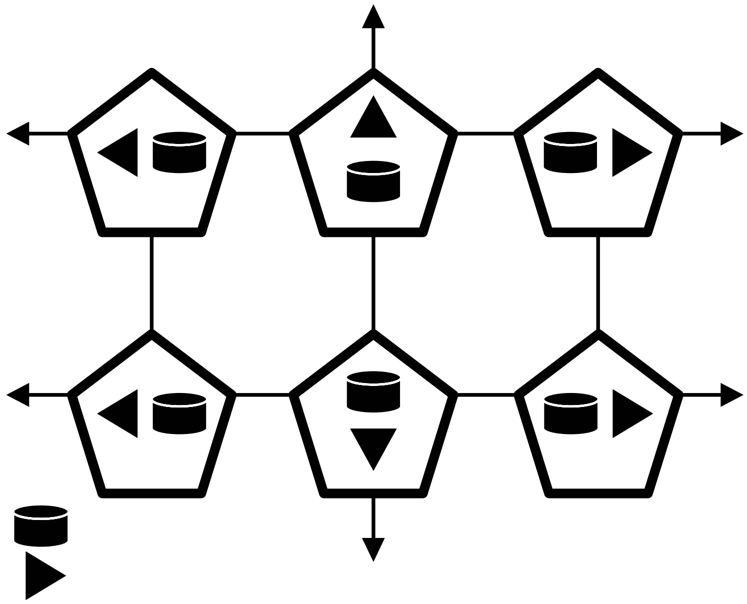

FIGURE 3. INTELLIGENT INTERCONNECT ARCHITECTURE

Ultimately, Intelligent Interconnect systems suffer from the same scalability problems as Basic systems albeit at a higher limit. Intelligent Interconnect systems also have the same disadvantages when storage requirements reach a threshold level that saturates the inter-block communication.

As with Basic Interconnect systems, Intelligent Interconnect systems "bundle" multiple functions into each VOD building block. To add only streams to an existing system requires the addition of building blocks that contain stream capacity AND storage. That addition is not coming for free.

VOD Systems with Independent Building Blocks

Distributed Component

The Distributed Component scheme divides the functions behind delivering on demand digital video into separate components. Whereas in the Interdependent schemes each "VOD building block" is a smaller version of the larger whole, a Component based system assigns specific tasks to specific devices.

For example, three basic functions are: Delivery, Ingest & Storage, and Management. A Distributed Component-based system using this functional breakout would call for three distinct component types: one that processes and sends out packets, another that handles the storage subsystem, and a third that provides overall management. Notice how the interconnect scheme shown in Figure 4 is separately scalable.

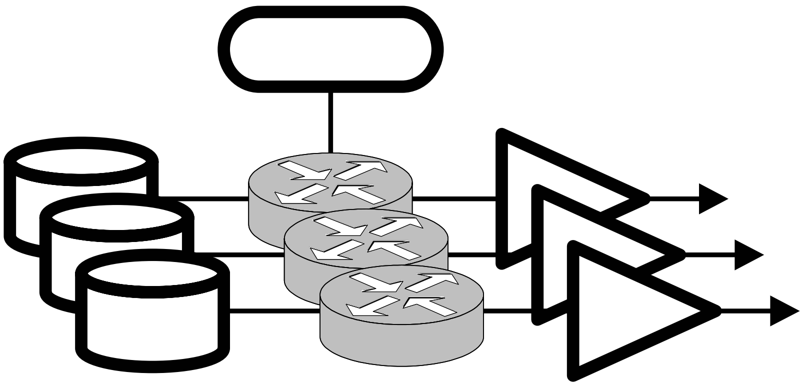

FIGURE 4. COMPONENT VOD ARCHITECTURE

When a stream request comes into the system, the management component assigns the stream to a specific delivery component. The delivery component determines the requirements for stream fulfillment and requests service from the storage component. After receiving a contract for service, the delivery component fulfills the stream request by processing and packetizing the media as it comes from the storage component.

There are numerous advantages for a component-based architecture. First, the functions can be divided according to whatever provides the best solution in terms of performance and cost. Second, each of the individual components can grow separately. This is perhaps its greatest asset. Additional streaming capacity can be added without adding more storage capacity, and vice versa. A component-based architecture is the most economical deployment scenario.

Third, redundancy is built-in because of the independence between components. If one of the delivery devices fails, then the management component simply assigns the stream request to an alternate delivery component. Fourth, the switching architecture (bandwidth) between the components can be expanded typically by adding more storage/streaming switch ports and extending the system. Lastly, a Distributed Component-based solution lends itself to open interfaces and best-of-breed component usage. Costs can be attacked at a functional level, for example, by changing one class of storage media for another. An array of fibre-channel hard disk drives could ultimately be replaced by an array of more economical Serial ATA disk drives sporting a fibre channel interface.

The downside with Distributed Component schemes is that some resource coordination/governance is required across component types. For example, a copy of a popular program held in one storage device may suddenly receive more requests than can be serviced. In order to preserve customer service levels, complete independence is not practical. The management component often provides this value-added governance.

A second disadvantage for Distributed Component solutions is the minimum footprint is often larger because at least one of each component is required. Whereas in an Interdependent system the minimum building block is often a single device, the building block for a Component-based system is at least one device for each component. Often Distributed Component solutions are not cost effective at the very low end (approximately 100 standard definition streams).

Intelligent Component

An Intelligent Component scheme refers to a Distributed Component-style solution wherein three optimizations are provided:

1. Each component is intelligent enough to help out others
2. Components can be unified onto a single platform
3. All components provide a failover strategy to maintain system operation

As discussed above, under Distributed Components, a delivery device can receive a contract from a storage device for stream fulfillment. The delivery component has intelligence added that provides a relatively small internal cache to be used to fulfill future requests for the same title. Therefore, the next request for that title requires no communication to the storage component. Taking this further, the delivery component can inform the management component whenever this case arises such that the management component shows an affinity between title and delivery component. Intelligent caching algorithms are required to utilize the memory effectively.

Intelligence about the network topology allows components to make better decisions on resource utilization. For example, new content propagating into the system can be placed into a specific location based on the predicted popularity. Perhaps a new release should be placed directly in delivery component cache or pushed out to an edge location. Such decisions can also be made dynamically based on the actual request patterns.

Another important attribute of Intelligent Component-based solutions is a platform independent architecture that permits each of the individual functions to be unified onto a single platform for small system deployment. This is particularly important for decentralized network topologies in order to maintain cost competitiveness and management efficiency of the solution.

Considerations in Scaling VOD

The preceding topics provided definitions and a survey of VOD solution architectures. To most effectively scale a VOD system requires a basic understanding of the underlying architecture. Equipped with relevant definitions and the pros/cons of various VOD system architectures, we will now discuss the basics of scaling VOD systems.

The Basics

All discussions about growing a VOD service begin with an assessment of the head-end basics. Depending on the deployment architecture, rack-space will be required at the Headend and/or Hub sites. Ideally, storage capacity upgrades can be placed lower in the rack due to the significant weight of disk arrays. Power requirements for additional storage and stream delivery hardware must be factored into the plan. Cooling requirements may be significant enough to warrant attention. Proper HVAC and power calculations should be made when contemplating any major upgrades.

The details of these attributes are outlined by the VOD server vendor, but generally speaking increased stream and storage *density* (streams/RU or storage/RU) provides relief in rack-space and reduces management burden. Next generation VOD server systems also typically call for lower power and cooling requirements than 1st generation equipment despite the increased density.

Another consideration at the basic level is a discussion about the long term strategy for the VOD deployment. Operators must anticipate the services they will offer, and the subscriber demand for those services, to develop a scalable system. Often initial deployments progressed with a decentralized architecture. ***Will that architecture best serve the needs of customers looking forward?*** By developing a three+ year plan for VOD services, the short term objectives can be balanced in the long term framework. Tremendous reductions in transport costs justify an evaluation of more centralized solutions.

One industry trend that will have a profound affect on VOD systems is the growing library of high definition (HD) content (defined as above 11 Mbps and typically 15 Mbps). HD content requires approximately four times the storage and streaming bandwidth as standard definition (SD, assumed as 3.75Mbps). ***Do you forecast HD content and if so, how much?***

Linked to the decisions about content encoding rates is QAM channel allocation logic utilized by the end-to-end network. Several operators are taking charge on this issue by providing an algorithm that VOD server manufacturers must support. Other industry players are seeking to move this decision outside the VOD server to a centralized system resource manager. In either case, the bottom line is ***What business rules do you wish to apply in the case of SD streaming versus HD streaming?***

For example, consider a 256QAM system in which each 6 MHz channel provides just over 38 Mbps of net throughput. This equates to ten SD streams. Consider the situation where 6 SD streams are already allocated to that QAM channel and the next request that comes in is for an HD title (15 Mbps stream). The decision to allocate the remaining bandwidth to one HD stream will now saturate that QAM channel. If in the next moment another 4 SD title requests come in, we have lost those revenue opportunities. Clearly, the value of the HD stream and the SD stream are not the same from a network and systems perspective.

A slight extension of this issue is scaling VOD to add sophisticated features such as Playlists. A Playlist is a set of titles that are put together and played in sequence to the customer. Playlists provide an easy mechanism for mixing ads with content on a more customized basis (quite different from ad insertion servers). The issue with Playlists comes in when the elements of the list are encoded with different bit rates. A sports program encoded at 5 Mbps could be interrupted periodically with 2.5 Mbps encoded commercials. Today's 1st generation VOD systems reserve enough downstream bandwidth for the playlist element with the highest encoding rate during the entire session. The best solution is for the VOD system to dynamically alter/negotiate the QAM channel bandwidth reservation.

Finally, as the VOD service grows it will require more efficient management and troubleshooting. ***How do you troubleshoot problems and determine root cause?*** With a greater number of building blocks in the network, the opportunities for congestion, contention, or timing problems grows. Operators require efficient logging and diagnostic systems that simplify the tedious task of combing through server log files by automating trouble alerts and providing meaningful error reporting.

Expanding Stream Capacity

As take rates increase (percentage of simultaneous usage), everyone is happy except perhaps for the VOD server engineers and network engineers that must expand the infrastructure.

An important first question to consider is ***What is the reason behind the increase in streams?*** The expansion of service by adding new service groups is quite different than service request denials due to QAM channel bandwidth saturation. In either case, unused stream capacity in another service group could be re-assigned (assuming a fully switched environment) but in the latter case the problem will not be resolved without node splitting or an increase in the number of narrowcast channels.

The physical steps in adding stream capacity will be unique to each vendor, of course, but the general considerations are the same. No matter what VOD system architecture (Interconnected or Independent), some additional gear (or software licenses) must be installed to provide the additional stream delivery functions. Similarly, the output switch (in the case of Gigabit Ethernet-based delivery) may need to be expanded. It is possible that an existing deployment will also require the addition of *storage bandwidth* when increasing streaming capacity. This is of course not possible in Interconnected systems with a saturated interconnection link – in those cases an entirely duplicate cluster must be purchased and installed.

The Streaming and Storage Relationship

Another attribute to evaluate within the VOD system is the relationship between storage and streaming. As discussed earlier regarding Interconnected systems, streaming and storage increase in tandem because each building block contains some of each. The operator pays for both even if only one is needed. These systems typically have very little caching intelligence in the stream delivery engine and therefore the performance comes down to a matter of calculating how many spindles (hard disk drives) are required. It is always OK to have more disk drives than necessary. However, the reverse is forbidden: to have less spindles than necessary for stream delivery will cause the streaming engine to be starved. This is similar to the garden hose analogy we referred to earlier—for a set level of water pressure (storage bandwidth), continuing to increase the number of taps on the garden hose (stream requests) will reduce the performance of each tap until the last added tap is worthless (data under-run causing a failed stream). Systems without robustness to protect against this condition would bring down the entire streaming capacity by starving packet delivery.

For example, a high-end 146GB fibre channel disk drive can sustain approximately 25 MBytes/sec (under normal, multi-file read operations). A set of eight disks would provide 200 MBytes/sec and hold 660 hours of standard definition content. Assuming content is striped across all eight disks and the request pattern is evenly distributed, this configuration could support in excess of 400 standard definition streams. Think of this as a "building block" that stores 660 hours of content and provides 400 streams. If you seek a system that provides 2,000 streams and 3,300 hours of content, you can simply assemble five of these blocks. But what if you have a system with 1,200 streams and 2,000 hours of content and you want to increase it to 1,600 streams? In this case, additional disks must be added to provide the additional stream support.

VOD systems with this **Interconnect** architecture spread the content out and replicate it to ensure enough storage bandwidth is present to support the additional stream count. This replication is part of the cost of expanding streams, yet shows up as additional storage capacity in the system.

VOD systems with a **Distributed Component** architecture handle scalability in a smoother fashion. Storage capacity can be added separately from storage bandwidth, and vice versa. So, to expand streams in a Component-based system, additional stream engine devices are added, and if necessary, the precise amount of storage bandwidth (storage interfaces). Note that it may be necessary to add spindles in some cases, depending upon the level of caching in the solution.

Consider further that **Intelligent Component** systems typically utilize next generation sophisticated tiered caching within the streaming engine itself to reduce the dependency on the storage component. By caching a popular title in hard disk or RAM within the stream delivery component itself, the stream request is fulfilled without utilizing the storage component. This sophistication can save upwards of 50% in the cost of storage in certain deployments and is worth considering as you evaluate how to scale your VOD systems. Understanding asset utilization patterns can optimize the caching strategy, reduce storage bandwidth, and eliminate over-provisioning of costly cache media (typically, RAM).

Considerations Beyond the VOD System

As streams are increased, ***Where/How will the requests be managed prior to reaching the VOD system itself?*** Typically backoffice software systems provide a common framework for server applications to enable customers to use specific services.

Can the backoffice software system handle increased transaction load? Stream start-up and stream completion (tear down) are particularly "heavy" on the backoffice system. Measurements should be taken to determine the peak transaction level to ensure it is not already running near its limits. Typically a backoffice software system upgrade requires significant maintenance time and disrupts numerous services.

***Can the digital network handle the increased load*?** Similar to the backoffice system, the network control system (DNCS, DAC, etc) must be able to handle the additional transaction loading. If session-based encryption is to be used, the load on these devices can be even greater. Several MSOs have measured the transaction performance of the S-A DNCS, for example, to peak out around 25 sessions per second under certain conditions. In a 5,000 stream system, what are the odds that more than 0.5% of customers will request a stream within one second of 8:00pm? A detailed understanding of your operating trends will indicate how close you may be to the limits of the current digital network control infrastructure.

***What expansions are necessary in the transport network*?** In centralized and hybrid deployment architectures, additional transport bandwidth is required when stream count expands. (The possibilities for expanding transport bandwidth are beyond the scope of this paper. However, it seems that centralized solutions are best served by a Gigabit Ethernet/DWDM infrastructure with either electrical or optical combining at the Hub.) Transport costs can be realized below $40 per stream in such cases. For operators with analog WDM or SONET/SDH transport, expansion is more costly. Hopefully for small increases in stream count, existing lit fiber can pick up the load.

Is the two-way plant (forward and return path) configured such that additional on-demand traffic can be handled? What other gear must be expanded? Any increase in streaming capacity must be matched with other equipment. QAM modulators and demodulators, RF upconverters, and multiplexers must also be expanded. This equipment will be located in the head-end or hub, depending on the deployment architecture.

In decentralized networks, the expansion of streams by adding service groups (another hub site) will require an expansion of storage. Care must be taken to propagate content to the new location before streaming can commence.

Finally, as mentioned earlier, the motivation behind the stream expansion is important. As you watch the service group utilization, the number of simultaneous streams for a given service group should not exceed the capacity. Commonplace allocation is four digital channels assigned to VOD, resulting in a maximum of 40 simultaneous standard definition streams per service group (256QAM environment). Any requests beyond 40 are denied translating to lost revenue, and perhaps more seriously, frustrate customers that we're trying to woo further into an "on demand" mindset.

The resolution for QAM channel saturation is straightforward: Either the node must be split, additional digital channels (spectrum) must be assigned to handle more streams, or a migration to switched broadcast architecture is required. A node split, which is best described as "halving the number of homes passed per service group" is often the easiest and lowest cost – estimates below $5,000 – yet depend on the HFC topology and availability of fiber. The reservation of more digital spectrum is usually the most difficult decision, particularly if the stream requests are seldom exceeding the available bandwidth. Digital channels (bandwidth) are in high demand across the various business segments of the operator. The conversion to a switched broadcast architecture is outside the scope of this paper but offers dynamic re-use of digital spectrum.

Expanding Storage (Titles)

The industry trend over the short period 2003 through 2004 emphasizes expansions in content offering rather than streams. It seems logical that investment in the size of the title library is necessary in order to drive increased take rates. The assessment of the quality of a VOD offering includes both the ability to get the specific title you want, but also on the intangible sense that you have a broad selection from which to choose. If any of us walked into our local video store and saw 500 copies of just the top 3 new releases, we wouldn't be happy—we want to know there is a plethora of titles.

Library sizes are measured in total hours of recorded content. In 2003, a library of 1,500 hours was typical. This year (2004), VOD offerings are expected to grow almost 50% into the 2,000-2,500 hour range. Content libraries are predicted to double again in 2005. Using standard definition encoding (3.75Mbps), an hour of raw content requires 1.6875 GB of storage. So at a minimal level, a 1,500 hour system last year required just over 2.5 TB which will grow to over 4 TB to hold a 2,400 hour library.

Note that the 1.6875 GB/Hour multiplier is just the starting point when determining how much storage is necessary. In fact, knowledge of the deployment architecture, the VOD system architecture, and the capabilities of the streaming engine itself must be considered to arrive at the actual storage requirement.

In a centralized deployment, expanding the library is perhaps the simplest: add storage media per the vendor's instruction. In a decentralized deployment, trucks may need to roll to hub sites to add capacity at each location. Hybrid deployments may require truck rolls for edge storage upgrades as well as modifications to the dynamic content caching algorithms. With decentralized or hybrid deployments, propagation bandwidth must increase along with content library increases.

Trick Modes and Storage

Within the VOD system itself, the amount of storage to add to cover the library growth may not be easily determined. Some VOD server vendors create so-called "trick tracks" for every original title. When first provisioned into the system, the original MPEG-2 file is processed several times to create a set of files to support each speed of fast forward and rewind a customer may request. So in fact, each title in the library may have four to six physical files associated with it. During trick play, the system switches dynamically from the original MPEG-2 play file to the appropriate trick track file corresponding to the speed/direction that the customer requests. The visual quality of these off-line generated trick tracks can be quite high, but a downside to them is the increased storage footprint required.

A good rule of thumb is to increase storage by at least 30% for these additional trick track files. Note that in a decentralized environment, the 30% increase applies to every streaming location. For a 2,500 hour library, using Fibre Channel or SCSI disks, the additional storage overhead translates to over $10,000 in raw storage capacity cost at each location.

Most next generation VOD systems dynamically process MPEG-2 "on-the-fly" to generate trick mode data streams and therefore do not have the high storage penalty or any significant delay in making new content available.

Striping Content and the Impact to Storage Growth

Clearly the total storage capacity in the VOD system must be greater than or equal to the library size. Systems that require separate trick tracks require additional overhead. Let's look closer within some VOD architectures and evaluate ***How does the VOD system store the content?***

Some VOD systems "stripe" content across as wide a range of the system as necessary. The motivation for system-level data striping is twofold:

1. Redundancy—no single building block contains all of one data set (title)
2. Performance—elimination of potential system bottlenecks from all streams requiring access to the same portion of storage ("hot spots").

System-level striping has merit, but the downside comes at storage expansion time. Consider a 1,500 hour system with 2.5 TB of content spread across all of the storage devices. To add another 1,000 hours to the system, the entire data set must be "re-balanced" across the 4.2TB of storage. If the entire library was not re-balanced, then redundancy would be reduced and performance would suffer.

The process of re-balancing is non-trivial. For VOD Systems it is required that re-balancing be performed while the system is operational. As a background task in the system, it could literally take weeks between the time that physical disks are added to the system and the additional capacity is ready for utilization.

Another consideration is the actual physical media used to expand the storage capacity. Historically, hard-drive manufacturers are able to double the capacity every 15-18 months. In 24-36 months, hard-drive models are discontinued or priced out of practical use. So a VOD system installed in mid-2002 using 73GB hard-drives must be expanded using 146GB or 292GB drives because the 73GB models are no longer available. It is crucial that the

VOD server vendor demonstrate how hard-drive support and storage capacity is provided over time to eliminate forklift storage upgrades.

VOD systems utilizing an Interconnect scheme often require that the larger capacity drives be re-formatted down to the size of the smallest in the set. This least common denominator approach enables operators to continue to add storage, but much of the capacity is not usable resulting in higher overall storage costs.

VOD Systems utilizing a Component scheme offer substantial flexibility in expansion of storage. These systems can utilize the sweet spot of the hard-drive cost/capacity curve over time. Since the storage component is separated from the rest of the system, it can be expanded intelligently with different capacities and bandwidths. The management component orchestrates the communication of component capabilities to fully optimize system performance.

Even further, Component-schemes can introduce entirely different classes of media within the solution. A system utilizing fibre-channel storage can be augmented by Serial ATA (SATA) drives with a fibre-channel front-end interface. Even Network Attached Storage (NAS) can be utilized. The limits depend only on the acceptable interfaces into the stream delivery components.

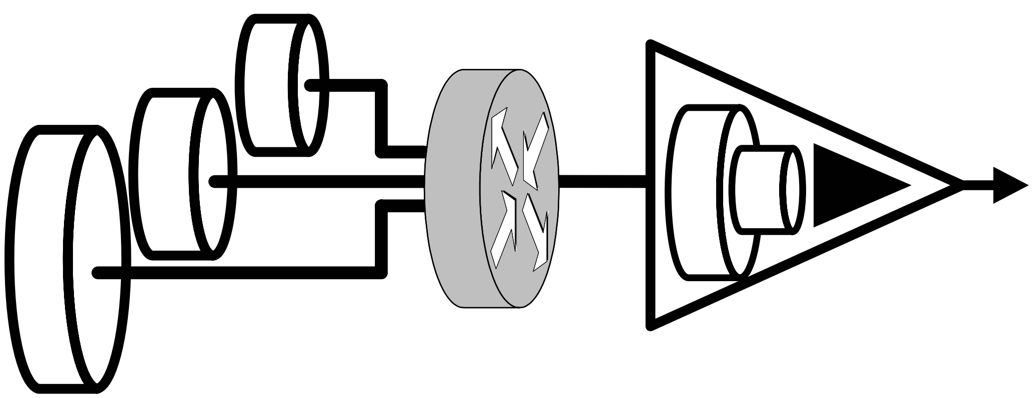

FIGURE 5. NEXT GENERATION TIERED STORAGE SOLUTION

As Figure 5 reveals, multiple tiers can be identified within the storage system and can be augmented by caching tiers within the streaming pumps directly. This tiered, or hierarchal, storage system optimizes both performance and cost. No Interconnect system deployed today offers this flexibility.

Content Propagation

Beyond the VOD system's internal architecture, further considerations remain about ***How is content propagated into the system and how will that scale with the additional library titles?*** Let's examine a few industry terms related to content.

The act of receiving content into the VOD system is called *ingest.* To *provision* a title is to enable it for play-out (service). In some systems, this means replication and re-encoding. In other systems, no further action is required. Total *propagation time* includes the effort to *ingest* and the effort to *provision.*

The industry has settled on two basic categories of ingest:

1. PULL. The content is "pulled" from some location by the VOD system.
2. PUSH. The content is "pushed" into the VOD system by some other device/mechanism.

PULL-based Propagation

Content propagation utilizing PULL can be as simple as a process within the VOD System that, upon command by the backoffice software system, checks a central storage location for new content files and then transfers the content into VOD storage (using FTP or other mechanism). This form of ingest is fairly straightforward because it is simply a network transfer that is highly tolerant of network bursts or storage prioritization and can be dynamically throttled by the VOD system to preserve performance of stream fulfillment.

The PULL method is in use within virtually 100% of VOD deployments today. A content receiving device (TVN Docking Station™ or N2BB MediaPath Catcher™ for example) holds the content after transmission from the content aggregator and until the VOD System has brought the title all the way to the provisioned state.

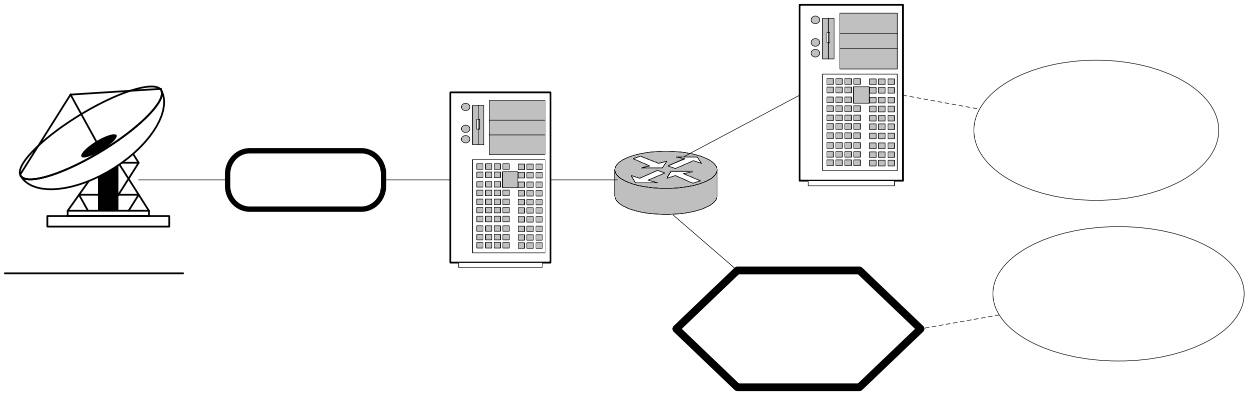

FIGURE 6. SIMPLIFIED CONTENT "PULL" TOPOLOGY

Content propagation time is of course a function of both the deployment architecture and the VOD system architecture. In centralized deployments, the title must be transferred once over the network and then processed internally by the VOD system. In decentralized and some hybrid deployments, content may be transferred numerous times to each streaming server location.

After the actual transfer time, the VOD system processing time to prepare the content for play-out can be extensive in systems that re-encode into multiple trick tracks. For every ten hours of content, an hour of processing time may be required.

The big picture question on content propagation as library titles are expanded is this: ***Is the propagation network sufficient for the increased load?*** For example, assume the content receiving device has a 100 Mbps interface that saturates at 80 Mbps. Assume further a library size of 1,500 hours, with 20% weekly refresh. In this case, 300 hours of new content must be propagated per week. Running at full speed, the 300 hours can be transferred to one location within the VOD system in 14 hours. The transfer can occur in smaller chunks of 2.5 hours during each of six daily maintenance windows.

Now consider expanding to 2,400 hours of titles, yet still maintaining the 20% weekly refresh. Now 480 hours of content will require 22.5 hours to transfer from the content receiver to the VOD system. The transfer can be broken down to a series of six 4 hour transfers during daily maintenance.

These numbers reveal very important scalability considerations for expanding storage:

1. Ensure sufficient network bandwidth is available for propagation. (Notice that the calculations above assume the transfer to just one other location.) If the VOD system must transfer the content over another network to remote locations, additional time/bandwidth is required.
2. If the VOD system must process and/or replicate the content within its architecture, additional time is required.
3. Transmission errors, or errors in the content itself that require another pitch, add complexity to what is already a time sensitive operation.
4. Content Ingest activity utilizes storage bandwidth as well and must be factored into the overall storage solution so as not to impact stream request fulfillment.

Other considerations for expanding the number of titles in the system encompass ***What is the performance of other network components?*** All content must be received by a content receiver device. ***What is the total available storage on the content receiver?*** When overlaying the schedule for receiving content during the week from the content aggregators, enough "working space" must exist to allow for smooth operation. It may be required to insert a separate Asset Management System (AMS) to improve content distribution.

The backoffice software system (also known generically as a business management system) is also affected by the increased number of titles. As the initiator of the content provision process and governor of problem resolution (e.g., failed transfer), the backoffice software must scale smoothly with the increased content management operations (refreshing a larger library each week, removing expired content promptly, etc.).

PUSH-based Propagation

Content propagation utilizing PUSH mechanisms is substantially different than PULL. We define PUSH as a transfer that is initiated by a device outside of the VOD System, and for which the VOD system must provide an absolute quality of service. The classic example of PUSH-style ingest is in Network Digital Video Recording (NDVR) environments: content is captured in real-time, shaped into constant bit rate (CBR) stream, and then transferred into the VOD System. In this case, the VOD system must guarantee the ingest performance. Only very low levels of buffering and re-transmission are possible in this real-time environment. More than a few errors will cause the captured content to be incomplete and most likely unusable. This can result in costly time spent by head-end technicians.

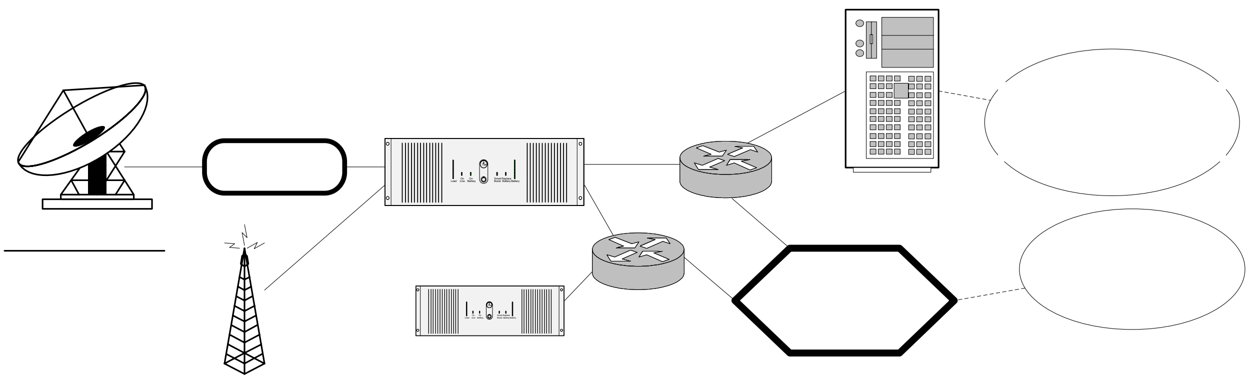

FIGURE 7. SIMPLIFIED CONTENT "PUSH" ARCHITECTURE

Whereas PULL content ingest tells the VOD System to "transfer as fast as you can whenever you are ready," the PUSH content ingest says "I am going to send this to you very slowly but constantly and you must give me attention the entire time." A sophisticated prioritization system must exist within the VOD system to accommodate PUSH-style ingest.

In terms of scaling VOD by adding more content, PULL content propagation is straightforward math. Increasing the bandwidth of the propagation network, or moving to an IP multicast form of distribution can resolve this potential bottleneck.

Much investment in research and technology development has been applied by some VOD server manufacturers to handle growing PUSH content. Interestingly, PUSH content does not place a heavy burden on the network in terms of bandwidth. Each PUSH stream is running in real-time, or 3.75 Mbps for standard definition content. A set of 20 channels requires 75 Mbps sustained. Notice also that PUSH ingest in a Network DVR sense has practical limits to scalability: Capturing every channel in the lineup (say 250 channels) would require one Gigabit Ethernet link. As recent industry events have revealed, content licensing issues continue to push out the timeline of seeing such large scalability with "broadcast capture" solutions. Therefore, utilization of PUSH ingest mechanisms is likely to be limited to a handful of channels for the foreseeable future.

Propagation Servers

Complex deployment architectures often benefit from the addition of a Propagation Server. This device performs the function of distributing content to all required locations within the

VOD system, thus reducing the load for the content receiver and AMS, improving distribution, and reducing technician time.

As content arrives, the VOD System is notified. The Propagation Server transfers the content from the content receiver/AMS, performs any content processing in a centralized fashion, and then manages distribution and replication according to the specific network constraints and streaming server requirements. The Propagation server is often a value added component in decentralized and hybrid environments.

As the VOD system scales up, content propagation requirements may dictate the addition of a Propagation Server in order to ensure content gets ready for play-out within the available time window. If one is already present, its distribution mechanisms may require upgrade. For example, an existing Propagation Server could be enhanced with IP multicast functionality to distribute the content to many locations using a single transmission.

Conclusion

In this document we have laid out a set of considerations to review while accommodating growth in a deployed VOD system. The same considerations apply to new VOD installations as well. Equipped with knowledge of the tradeoffs of various VOD system architectures, operators can develop a long term plan with confidence. The addition of streaming and/or storage capacity is rarely as simple as plugging in a new device and calling it complete.

Operators who understand the scalability criteria for streaming, storage, propagation, and ingest will be better equipped to make architectural decisions. Better decisions in deploying a VOD network will result in that network being able to scale with the service growth and not require replacement.

About the Author

Mark Jancola is the Vice President of Advanced Products & Technology at MidStream Technologies, Inc, an innovative leader in next generation Video-On-Demand (VOD) and Network DVR (NDVR) solutions. Over the past two years, MidStream led the VOD server industry in helping to deliver the first NDVR deployment in a cable system. MidStream's development efforts remain focused on technologies that will improve the competitive edge in the services operators deliver. Mr. Jancola has more than 15 years of engineering and technical experience building multimedia platforms and solutions. Prior to joining MidStream Technologies, Mr. Jancola was a member of the Cisco Systems Internet Communications Software Group, and earlier, was director of product development at Active Voice Corp., a leader in enterprise-based communications and unified-messaging servers. Mr. Jancola holds an MBA and BSEE from the University of Washington and is a member of IEEE and SCTE. He can be reached at mark@midstream.com.