Content Management Systems versus Content Delivery Networks

Michael Adams
TANDBERG Television

Abstract

In today's on-demand deployments, content management and content delivery are intertwined into a store-and-forward, multicast delivery system that relies on satellite transmission for secure asset delivery. The mode of delivery is a "push" from the content provider to the on-demand site, such that all assets must be provisioned in the video-server memory before they are offered to the customer. This same system ensures that titles are offered in a consistent and flexible way according to the operator's marketing requirements.

In contrast, cable's high-speed data services use content delivery networks (CDN) to deliver web assets, as a series of file-transfers over terrestrial networks, such as leased interconnects and backbone networks. Moreover, the mode of delivery is a "pull" from the subscriber, based on HTML requests from the website. Websites are essentially unmanaged by the operator, except for caching frequently served pages to manage backbone bandwidth.

There is considerable interest among cable operators in migration towards using a CDN approach for on-demand content. However the transition is not trivial because of cable's unique requirements related to scale, management, and security for on-demand assets.

This paper explains in detail the requirements for a Content Management System (CMS) to manage end-to-end delivery of on-demand assets in a hybrid network environment using satellite and terrestrial links. We show how a migration towards terrestrial content distribution is possible without giving up some of the fundamental advantages of today's "push" delivery model, namely that very popular titles do not generate unmanageable spikes in network utilization and that customers are never offered "hot" assets until they have been successfully provisioned at the video-server.

Nevertheless, there are excellent reasons to migrate to terrestrial networks for content distribution including the ability to turn-around assets much faster when required; an example would be a political campaign advertisement. In addition, the explosion in content choices, coupled with the increase in the number of high definition titles, makes the "push" content delivery model impractical at some point. Therefore, mechanisms to identify certain, less popular titles as "library content" become essential, as does selectively employing a "pull" content delivery model.

Finally, this paper will contrast the requirements for a CMS with those for a CDN, and show how the two technologies can be used together in certain circumstances to blend the advantages of both types of approach.

INTRODUCTION

Video-on-demand systems have now been deployed by all major cable operators across their entire footprint and these systems work extremely well, serving millions of customers on a daily basis.

Cable on-demand services are based on a QoS guaranteed, connection-oriented model from the streaming service to the subscriber. This model relies upon session resource management which allocates a guaranteed slice of bandwidth between the server and the set-top for the duration of an on-demand session. This is possible because the path from the streaming server to the set-top is over a relatively simple access network topology (typically via a Gigabit network to a QAM modulator). The only possibility of blocking is in the HFC network itself, but this well-managed by allocating constant bit-rate sessions using standardized QAM resource management. Cable operators have become adept at increasing the total bandwidth per subscriber by reducing the size of the service group as on-demand peak usage increases.

However, this means that video servers have to be placed at the edge of the network and they need to be pre-provisioned with all the assets that the customer could possibly want. This leads to some challenges to solve in deployed systems:

1) Scaling the number of assets from 10,000 to over 100,000 to support long-tail content.
2) Effectively managing a very large number of assets across multiple sites.
3) Ability to dynamically change metadata independently of content propagation.

In this paper we will examine the role of a Content Management System (CMS) and a Content Delivery Network (CDN) in solving these problems and show how they address different aspects of these challenges.

ON-DEMAND CONTENT DISTRIBUTION

Content distribution products were first developed to meet the specific needs created by cable operators as they deployed on-demand systems on a headend-by-headend basis.

In the earliest days, content files were shipped around on tape and metadata was manually entered into the VOD system at each headend. In order to scale on-demand services, the concept of a package was developed – this is essentially a collection of assets and metadata files that completely describe the on-demand title. Each package is transformed into a single file using a UNIX utility called tar. These tar files are transmitted over satellite, using multicast to greatly reduce bandwidth requirements. Thousands of hours of content are distributed to systems via satellite from multiple content providers. Although an hour of standard definition content requires about 1.7 gigabytes to be transferred, the actual transmission rate from each content provider is relatively low, in the order of 10 Mbps.

Satellite distribution provides an extremely efficient, "push" mechanism to get the required content to a very broadly scattered set of cable systems. This so-called "pitcher-catcher" approach uses a reliable multicast algorithm to make efficient use of satellite capacity to push assets to a large population of headend systems simultaneously. Also included is robust encryption to prevent unauthorized access to the content files.

Unfortunately bundling the content files with the metadata means that the metadata is also multicast, leading to a one-size-fits-all approach. This forces the operator to make metadata changes at each system, after content distribution rather than centrally. As a result, asset management has become complicated and time consuming because each system is managed separately. Worse still, there is a lack of transparency to corporate marketing and central operations groups. If delivery of title to a site is unsuccessful, a manual "re-pitch" is required, which is both costly and inefficient.

The title metadata includes fields on pricing, categorization, and availability window. A content provider distributes packages with these values set according to its business rules. The operator may need to override some of them such as different pricing on a system-wide or system-by-system basis. Some assets may need to be filtered because they are not required at the particular system, or because there is insufficient storage capacity for them. Because of the need to change the metadata for each operator, and even down to the specific site, “Asset Management Systems” were introduced. Initially asset management systems were deployed on a site-by-site basis but they have evolved to support multiple sites in a regional environment.

CONTENT MANAGEMENT SYSTEMS

On-demand systems rely on thousands of hours of content distributed via satellite from multiple content providers. With so many files in so many locations, it becomes a complex task to keep track of everything and today’s manual systems are struggling to keep up. There is often a significant cost in on-demand operations due to so-called “content propagation errors”. These occur when part of the content distribution chain fails during the provisioning process. In most cases, the harm is done because there is no mechanism for the operator to automatically check that all the necessary assets have been placed on all the appropriate servers. So the provisioning error is often discovered first by the customer, who requests a title and gets back an error code. This impacts the session success rate. Worse, it discourages the customer from using the on-demand service and reduces operator revenues.

In practice it is very difficult to make the content distribution chain completely “bullet-proof” and doing so would come at a significant cost, so it is better to automatically check the asset status on a regular basis. If a particular asset is missing from a server, it can be automatically rescheduled for delivery. Since assets are typically propagated before they are made available (before the viewing window), it is possible to repair any missing assets before they have any customer impacting effect.

The situation is similar for on-demand dynamic ad placement – in this case it is especially important to verify that the ad content is actually on the server before it is scheduled to be inserted into an on-demand stream. The ad decision is usually made at session start-up time, and at this point a play list is presented to the server.

Since operators offer the majority of on-demand titles across their entire footprint, it makes sense to manage them centrally. Since content distribution and delivery is spread geographically across a large number of systems, a distributed content management solution is required with centralized control. Since local divisions and regions also want to offer local on-demand titles, the solution must support local management of those titles.

A central control panel to manage content metadata provides a long list of advantages for the operator, including the potential to enable powerful and timely promotional campaigns along with pricing discounts, to provide targeted advertising support, to adjust the viewing window, to remedy errors, or to enrich metadata after the title has been provisioned (for example to support extended descriptions, advanced search, or recommendations).

Let’s explore some example scenarios that are problematic for operators today and I’ll explain how a content management solution can help to make these scenarios much easier to manage effectively. I’m going to focus on scenarios that are operationally intensive or impossible to implement today:

The corporate marketing and programming department believes that a title has truly underperformed and decides that a promotional campaign along with a price reduction is justified in the final week of availability. With centralized control of the metadata and the resulting ability to easily lower the price for the title, the marketing and programming department now has the ability to pull together a cross-channel advertising campaign in partnership with the content provider. Further, they can use their content processing solution to embed ETV triggers in that advertising, to direct viewers directly to the "buy" screen.

A set of titles with a common element or theme can be listed in a special category – for example, all the movies featuring "Ben Kingsley" or all the "Bond" movies. By cross-linking the metadata of titles in this way, this also creates a foundation for automatically generated recommendations for new titles.

An unpopular title can be removed from the system before the end of its availability window to free up storage space for new titles. As new titles are introduced, a set of rules is used to distribute them to each market according to pre-defined priorities so that the on-demand storage in each market is optimized for maximum revenue generation.

A set of titles can be ingested in the traditional on-demand infrastructure, and then automatically processed according to a set of rights and rules housed in a seamlessly connected solution to the appropriate content compression and metadata formats for broadband and mobile video service platforms. Depending on the rights and rules, in some cases, only the trailers are sent to these additional platforms, and in other cases, the movies can be purchased and viewed on them.

In order to do all of the above effectively, a fundamental change is required in the way that the content management solution structures and handles the relationship between a so-called "heavy" asset and its associated metadata. In short, the processing, managing and delivering of metadata needs to be separate from, but not disconnected from, the associated assets, and vice versa. The content management solution needs to maintain more intelligence, awareness and flexibility in this critical linkage between assets and metadata.

For example, for certain distribution needs, the delivery of metadata might be unbundled from the delivery of the associated heavy asset. This is the case for many operators that do not have the backbone capacity to handle distribution of their content asset. By separating the delivery of these two assets, the operator can deliver very large content assets to local systems over a very cost effective satellite network, and yet still deliver metadata via their intranet. Thus metadata and rules for the management of the metadata can be organized centrally and delivered to the local systems whenever content is distributed or updated.

Finally, operators will need to be able to provision content so that it can be played on any device. This provides the operator with a competitive response to "over-the-top" video providers by providing a seamless extension of on-demand cable services to additional devices within the home; for example, to TVs with broadband connection, to PCs, or to an ultra low cost IP set-top box. To make multi-platform services manageable and scalable, the content management solution must enable the operator to provision a single title with multiple playout options across multiple devices.

CONTENT DELIVERY NETWORKS

Content Delivery Networks have grown up to support web-based delivery of information, including arbitrary content formats for graphics, music and video. Wikipedia defines a content delivery network as "a system of computers and storage networked together across the Internet that co-operate transparently to deliver content most often for the purpose of improving performance, scalability and cost efficiency to end users".

Essentially a CDN is a cache-based system where files are moved around the network based on the pattern of usage. Considerable ingenuity has been applied to the algorithms and protocols to do this. Peer-to-peer (P2P) protocols are one example of an approach to building a distributed, collaborative network of content stores.

A typical CDN implementation is to populate files into an origin server (which itself may be a distributed entity). This server is connected to many cache servers by network links. The cache servers directly serve the clients, which are web-browsers. When the user navigates to a web-page, the embedded markup language (HTML) causes the browser to request all the necessary files for that page using HTTP requests. In the first fetch, the cache server must retrieve them from the origin server and pass them along to the client, keeping a copy in the cache. On subsequent fetches, the local cached copy can be used to save bandwidth on the network link back to the origin server and to save processor bandwidth on the origin server.

The CDN architecture allows for multiple tiers of cache servers for better scalability and allows the edge cache server to be pushed closer to the client, allowing for faster response times and higher bandwidth across a given link. However, because each client request is independent of every other, each fetch is a unicast session and considerable effort has been applied to cache management algorithms, load-balancing, and so on to make best use of each fetch operation.

Although this is a simplified overview of how a CDN works, it captures the important principles of operation. It is fair to state that CDN technology has been developed for the general case of managing a large number of small files with little regard to what function those files actually represent at the system level. As such, the model lacks a higher level abstraction such as the asset grouping and its metadata that ties certain files together. The reason for this is that the hyper-linked web pages themselves provide the framework and the navigation for the website. Essentially the "metadata" equivalent is embedded in the HTML markup of the web pages. Web technology improvements including dynamic HTML, applets, and servlets make web-sites much more dynamic and flexible than before and allow support for localization and personalization, but this principle remains unchanged.

Although the CDN automatically responds to client requests to make ensure that the most recently used files are held in the cache servers to satisfy overall demand, this may not provide the best indication of asset popularity. This is because many cached files are related to navigation requests versus asset requests. Moreover, there is no intelligence in the CDN to pre-position files in anticipation of a demand for them which may cause unpredictable spikes in bandwidth as new, popular assets are discovered by the clients. [1]

CMS VS CDN

Table 1 summarizes the functions supported by CMS and CDN technologies.

Function	**CMS**	**CDN**
Metadata management	Y	N
Transcoding control	Y	N
Workflow	Y	N
Content Distribution	N	Y
Multicast Distribution	N	N
Cache Management	N	Y
Load Balancing	N	Y

Therefore, we conclude that CMS and CDN functions are complementary to each other; a CMS allows for functions related to the semantics of the content while a CDN provides the mechanism to distribute the various files across the network of servers.

The next step in this analysis is to discuss how the two technologies can best be combined in a practical way.

CMS AND CDN

In this section we will discuss a potential migration strategy to add CMS and CDN technologies to an existing on-demand deployment. As we have already discussed, CMS can provide a powerful management layer that will work with the existing satellite content delivery model to provide operational advantages and additional control and monitoring capabilities.

Therefore, the first logical step is to add CMS to the existing on-demand deployment to reap the benefits of this technology.

As demand for long-tail content increases, the system may reach a limit on the number of assets that can be "pushed" to the on-demand systems in the available time. Ultimately this is limited by the cost of satellite bandwidth available to content providers and the cost of storage in each on-demand system. Titles towards the end of the long-tail distribution will be so rarely viewed that it does not make economic sense to distribute them and store many copies of them locally.

At this point, CDN technology provides a complementary addition to the existing "push" content distribution model. It should be noted that the latest generation of "pitcher-catcher" based solutions will operate across a mix of satellite and terrestrial facilities. Nevertheless, for popular titles that are expected to generate significantly more than one play on average per system per month, the "push" model is still superior from a cost perspective because of its use of a multicast distribution protocol. Therefore the best solution is to use the CMS to implement distribution rules based on popularity indicators embedded in the metadata in order to select which delivery mechanism to use on an asset by asset basis.

CONCLUSIONS

In this paper we have reviewed the capabilities of CMS and CDN technologies in on-demand video deployments and conclude that:

1) They are complementary technologies and there are advantages to using the two in conjunction when the number of assets is greater than 10,000.
2) CMS technology is a useful addition to existing on-demand deployments in conjunction with todays "push" content distribution systems.
3) CDN technology may be added as an additional distribution technology for long-tail assets, essentially "pulling" these assets from a central content library only when they are required on the first customer play.
4) Extensions to support multi-platform delivery of assets will require a CMS platform to manage the additional complexities of content transcoding and multiple content files formats per asset.

---

[1] ***THE BENEFITS AND CHALLENGES OF DEPLOYING LARGE REGIONAL VOD ASSET LIBRARIES by Michael W. Pasquinilli, Sunil Nakrani, Jaya Devabhaktuni, 2008 NCTA Technical Papers™***