A look at the ways that the User Interface and Video – Live and On-Demand - can benefit from a more cloud based architecture

A Technical Paper prepared for the Society of Cable Telecommunications Engineers
By

Robin Wilson
VP Business Development
Nagra
Robin.wilson@nagra.com

Overview

This paper cover the potential impact of cloud based services for cable operators on User Interfaces (UI's) and Video Encoding– both Live and On-Demand.

There may be a strong business case for "Cloud" where applications have large peak loads compared to the average load. Here the business case for outsourcing to a cloud maybe much stronger. And applications where the processing required is relatively high compared to the amount of data transfer required are particularly well suited to being Cloud based.

We look at the trade-offs between hosting parts or the entire user interface (often referred to as the UI) in cloud-based servers. There are many technical tradeoffs in terms of responsiveness, robustness, bandwidth consumed and flexibility. The optimum in most systems using STB is likely to be a hybrid arrangement where some UI elements are locally cached and some aspect of the UI is sourced from the cloud. For devices like tablets, it may be more efficient to source most if not all of the UI from the cloud.

The Audio and Video layers also have challenges related to the huge bandwidths and yet near perfect reliability demanded. Video encoding of live content requires high bandwidth, high reliability signals in both directions to and from the cloud. In many use cases this will make cloud based live video head-ends a challenge. However for some smaller scaled OTT offerings with little focus on live content, cloud basing may make sense today.

The Cloud Definition Morphs Over Time

The term "Cloud" used in communications goes back at least 50 years. Some of the earlier definitions covered telephony, Asynchronous Transfer Mode (ATM) networks and then IP networks. There are even HFC Cable networks reduced to a cloud diagram. In these early cases the rough definition was along the lines of "trust us, the signal goes in to this blob and magically finds the right way out – you don't have to know any more". To many these early definitions are still valid – a cloud can be just about any IP network. In this context the cloud is relatively passive other than a switching and routing function. There is not real processing or computing going on beyond these functions.

The more recent definitions (around last 8 years?) encompass extremely flexible efficient general purpose computing platforms that run virtualized processes that can be managed and billed in a very efficient and agile manner. These platforms are particularly attractive where processing needs are highly variable. Without this most recent cloud concept, those needing large bursts of occasional computing power had to scale their servers for the peak load and potentially leave huge investments idle for large amounts of time. Terms like Software as a Service (SaaS) or Platform as a Service (PaaS) are often used to describe a relatively new form of technology and a new business model.

The ability to be highly agile makes cloud computing very attractive to enterprise users. For example, applications run on the cloud are often highly asymmetric client-server applications with the client being a simple and largely standardized web browser. This allows very rapid updates as only the highly controlled server image has to be updated.

As cloud technology has developed and has become much more powerful, it is now capable of handling User Interface (UI) processes and even some video processes applicable to Cable.

Cloudwashing

The term "cloud" most recently has suffered from a great deal of ambiguity and hype. Indeed the term "cloud-washing" has been invented to describe marketing based attempts to reposition traditional client-server applications as being cloud based. In many ways the true definition of cloud is in the mind of the beholder. Clouds are nothing more than racks of rentable servers or virtualized processes running on these servers. In many cases the applications running on a cloud based server could equally well be run in a private cloud or data center operated by a cable MSO, or even on a server located in a head-end rack. So to a large degree the term cloud is all about business optimization, flexibility, and virtualization technology and less about marketing hype.

HTML5 User Interfaces

HTML5 is often presented as a wonderful future proof solution to rendering both the navigation elements of the UI and the underlying video stream (*The HTML5 Opportunity - Building Seamless Cross-Device User Experiences, Huntington, Wilson, SCTE 2012*).
While HTML5 has many positive attributes, and many more still under standards development, it is at its heart the client of a client-server relationship. The client-server aspect of HTML5 lends itself well to cloud based server architecture for a STB, PC or Tablet based client. However the words "HTML5" and "Cloud" are often wrongly assumed to be inextricably interlinked. While the server (web server) for the HTML5 browser client could clearly exist in a cloud, there are many reasons why it often does not. Part of the reason for this common assumption is that web browsing assumes the internet and that the internet rightly assumes a cloud location.

If we look at STB applications we need to take account of the bandwidth and reliability of path to and from the cloud. It is unlikely that any even marginal impairment on the availability of the video path can be accepted. If for example the data path over DOCSIS was down or impaired, a purely cloud based HTML5 server for the UI may fail and inhibit any form of navigation or even basic functions like channel change. Because of this, it may be a more risk free strategy to serve some of the UI locally from the STB. This is not unlike many consumer "appliances" like Wi-Fi routers that have their local web server for basic UI functions. In addition, this local strategy allows the STB to cache what could be days or even weeks or rich guide data, saving bandwidth and potential latency issues. Thus the local HTML5 web-server is set up to cache the data and can intelligently and efficient refresh as needed. The cache serves as the fall back data repository during failure or impaired service.

With this cached model it is still possible to further optimize the serving of the web or HTML5 elements to the browser. This is where the cloud may play another role. For example, rich meta data for the guide or navigation applications may contain rich graphics and dynamic advertisements. These could even be video streams. Here a remote cloud or a local cloud may be a perfectly acceptable source to serve these streams and images while the guide data still cached locally.

Remote User Interfaces

Remote user interfaces are often assumed and loudly marketed to be cloud based. In many cases the truth is a little different. Just as HTLM5 does not have to mean exclusively cloud, remote UI's do not either.

RVU and CCP-2 and CCP-3 are either deployed or in-progress standards that enable "thin client" remote user interfaces on low processing power STBs or even directly rendered in some TV's. While in theory the UI could be sourced from afar from a cloud, the reality is that a UI server in the home gateway is the most practical use case (and so far the only use case). Here the cloud description is stretching the concept of remote rather far – the remoteness may be just 20 ft.

Another clever tactic that has been developed, is to remotely render the UI in a server as an encoded MPEG stream. Thus there is no need for graphics processing or overlay within the STB. This may be ideal for remotely rendering rich UI's including HTML5 into very low powered legacy STB's. Again while the servers responsible for this rendering could well be located in a cloud the most likely commercial and technical solution is a bank of dedicated servers in a head-end – a form of private cloud.

Video Processing and Encoding

Here the applicability to being cloud based depends very much on the exact use case. Video encoders and transcoders are now widely available as software products that can be virtualized to efficiently run on standard cloud based servers though some require slightly specialized GPU acceleration capable servers. This is a fairly recent development in the last few years. With the advent of more compute intensive HEVC encoding, more and more encoding products will be compatible with cloud platforms.

There are two main factors that influence the technical and business sense of using a cloud:

1) Peak versus Average Loading

As described in the introductory section, the cloud business model is very advantageous when the encoding needs are highly variable and particularly with high peaks and low average.

2) Bits in and Out

Video applications can have huge bandwidth needs. The costs of transporting all these bits to and from the cloud may outweigh any gains achieved through other means. Here, scale may play a major factor. For example, a file that is uploaded once and transcoding to a limited number of streams may make business sense if it can be transferred to a CDN just once. If the stream is live for example and is by definitions used once, that may become prohibitive in bandwidth transfer costs. And a live uncompressed input signal maybe highly technically and cost challenging particularly for higher resolutions.

CDN's as a Cloud?

Some consider CDN's to be either a cloud application or "in the cloud". While CDN's may have some of the apparent attributes of the more recent cloud definitions, they lack some aspects of flexibility in use. In some ways the CDN application is dedicated to the hardware underlying it unlike a true cloud application. A further proof point is that some of the largest cloud computing providers outsource their own video operations to well-known pure play CDN's.

Regardless of the semantics, CDN's do have the cloud like positive attributes that they can be scaled for the user and dynamically billed based on usage or bandwidth metrics. And CDN's are critical for video delivered using other cloud based processes including OTT applications.

OTT and the Cloud

OTT is a combination of several potential cloud based applications. UI's, CDN's and Transcoding all have been covered above. For OTT where the viewing device is a tablet or PC, clearly the UI makes most sense to be cloud based. And as state above, CDN's are already somewhat a closed cloud that even the largest cloud operators prefer not to emulate. The video layer is more challenging but some aspects of transcoding may make sense in the cloud. Related to transcoding, there are also the packaging and encryption processes that also have the potential of being cloud based depending on the scale and peak to average load factors mentioned above.

Latency

When used in cable, cloud based applications typically still run over the DOCSIS path in an HFC plant. Here the latency is well understood and can be managed to keep round trip delays to be well under 50 ms. For UI applications, this delay may be just perceptible to a gamer. To a typical user this would likely be imperceptible.

Cloud based applications have two additional delay sources – the server process itself and the path to and from the cloud.

The cloud server should be capable of running any processes at least as quickly as if the processes were running on other servers. Indeed, if the process was run locally on the STB for example, the latency might even be higher. So in this context, cloud hosting should not hurt and potentially help reduce latency.

The path to and from the cloud servers from the cable head-end can be a major variable and vulnerability. This is where major differences can occur. For example, if a private cloud is used, it may even be in the same building and certainly on a carefully managed and controlled network. If a “public” or commercial cloud is selected, the quality of service on the interconnect is critical. This also becomes a potential risk point in the reliability and performance of the system. The potential delay here incudes a much smaller prorogation delay derived from the speed of light and based on the speed of propagation through the physical network (fiber). Worse case, coast to coast this particular delay should be less than 30 ms.

For some processes like billing and customer service, long delays may be acceptable. For UI applications, every 10 ms is critical. Here the path to the cloud can probably run over a well-managed commercial backbone. For video, even single ms can make a difference particularly at higher bit rates and where constant re-negotiation might be required such as in ABR (Adaptive Bit Rate) and OTT applications. In this instance a great deal of care needs to be taken and it may be that it is just too inefficient to run video to the cloud. (Other issues like the cost of bandwidth may also seriously disadvantage cloud applications for many forms of video.)

Conclusion

As the processes and applications that used to be performed by dedicated hardware or server platforms become more and more independent from the underlying hardware, using a cloud platform becomes a more viable option. While the processes and applications used in Cable may not have the same attributes as those used in enterprise, some aspects may lend themselves to cloud application today. In deciding what may make most sense it is important to consider the tradeoff between an out-sourced cloud, a private cloud and a local server farm. This ultimate decision is a complex one depending on many factors including scale, security, latency, peak to average ratios, bandwidth costs and how virtual the application can be.

Through all the recent cloud hype a bright ray of reality is starting to shine through.

Abbreviations and Acronyms

ABR	Adaptive Bit Rate
ATM	Asynchronous Transfer Mode
CDN	Content Delivery Network
GPU	Graphical Processing Unit
HEVC	High Efficiency Video Encoding
IaaS	Infrastructure as a Service
IP	Internet Protocol
MSO	Multiple System Operator
PaaS	Platform as a Service
OTT	Over The Top
SaaS	Software as a Service
STB	Set Top Box
UI	User Interface