The Value of High Endurance Flash

The successful delivery of a cloud-based DVR/PVR (Digital or Personal Video Recorder) product must focus on many critical issues that include reliability, cost and product quality. Real-world experience in delivering cloud DVR has shown that the product must provide a very close approximation of what consumers have come to expect from local DVR.

Users are likely to experience differences between cloud DVR and local DVR products when they use "trick modes". Fast forward, rewind and pause are the most important of these. Edge-located non-volatile memory can help fill this quality gap.

So what are the underlying causes of these trick mode differences in a conventional QAM-based cable television plant?

"That Depends" will probably be the answer you will get.

First of all we need to get a description of what it is that really causes the user dissatisfaction. These are divided into two important areas:

1) **Predictability of controls.** These types of problems are usually based around issues where a user presses a control or activates a feature and then has to wait for either an unpredictable or over-long period of time to get a response from the cloud DVR systems. We will discuss these issues below under **User Interface Latency**.

2) **Accuracy of “landing points”**. These types of problems are typically displayed as undesirable time offsets when using trick modes — for instance — press "pause" at Time A and see paused content from Time A + x

Let's discuss these under **Indexing & Caching Accuracy**.

The trickiest types of these various impairments occur when switching between live broadcast events and buffered copies of these events in order to simulate DVR pause buffering. These are typically called "Live Replay" tasks.

User Interface Latency

Defining sources of latency for cloud DVR in a QAM-based cable television plant can be very complex. Any networked DVR/PVR solution can experience unwanted latencies in backbones (up to an edge delivery server/CDN); in transmission (typically from backbone latency and edge server congestion) and at internal device software (timeouts/unreceived acknowledgments). However, many sources of latency are well tolerated by a typical user, such as during initial program starts.

Let's focus only on latency issues that are likely to negatively influence the customer experience. The most important of these occurs when a content stream is changed, as would occur when changing between streams with different content delivery rates (real-time vs. fast forward/rewind) or when switching between fixed frame displays and content streams (as when selecting or releasing a pause event).

Typical human perception of event thresholds occur between 10–25 milliseconds. Most people can readily tolerate event transition delays of up to 500 milliseconds if they occur predictably. However, it is important to keep DVR action response times under 250 milliseconds in order to meet user quality expectations. These thresholds can be very hard to meet if all Live Replay events have to be sourced from cloud service locations. Industry experience shows that cumulative delays caused by network device hardware, home-to-headend drop networks, CMTS systems, backbone networks, edge caches, video servers, business process systems and other systems can easily exceed 500 milliseconds, with a range that can exceed 5 seconds in high traffic conditions.
Indexing & Caching Accuracy

Probably the most frequent complaint about cloud DVR services comes from issues with accurately calculating “landing points” from data caches located in a cloud server system. You might have seen these types of issues yourselves when you pause a cloud-based service and then resume, only to have the content index jump suddenly to a different point in a stored program. This can also be seen when using trick modes in Live Replay or recorded content to skip past commercials. When you hit “Play” to get back to real time playback, you find yourself in the wrong part of the program and have to go back to trick modes to regain the correct program position.

The calculation of landing points from caching is an issue that is as least as complex as the determination of points of latency in a complex network in real time.

Here are some points to consider:

- Any caches of digital video are not continuous. Standard encoded video is compressed using different types of data frames and any replay must first be time synchronized before it can be displayed. As such, the more accurate your indexing is, the less time is needed to get a stream output to start and the more likely that your landing points are accurate. Indexing and stream compression are related, as there are fewer accurate landing points (i-frames) in highly compressed video.

- Live Replay time shift content is demanded without prior planning. Locating caches at system edges makes it possible to record all content without restrictions normally found in headend stored data licenses, such as mandate for user-initiation per session.

- Edge storage can take advantage of high quality hardware-based indexing that is available in all modern video SOCs / CPUs.

Real World Experience

So let's look at some real world issues that have been encountered when implementing Live Replay using Flash memory at a system edge, in conjunction with cloud-based DVR services. This service architecture is now feasible due to the availability of high endurance Flash memory, and has already been deployed in sufficient volume worldwide to have a well proven track record.

Using flash memory for Live Replay caching has some very positive effects on Landing Point accuracy and elimination of latency. Local hardware-enabled playback has latencies measurable in micro seconds, not milliseconds so there is no perceptible time between user action and device response.

There are other hidden advantages of edge storage for DVR caches:

- Content that is played back from centralized cloud sources needs to be encrypted using the same baseline QAM-based encryption that is used for live broadcast and/or VOD. These types of broadcast encryption systems can add multi-second start-up latencies that can be easily observed when starting up typical prerecorded content streams. In addition, calculating the real position of stream indexes on encrypted sources can also add additional jitter imprecision which needs to be considered when designing a content buffer or playback system. These issues do not apply to content stored in Flash edge caches, which typically use alternate methods for content protection.

- Locally cached data can be stored and retrieved much more efficiently when compared to cloud storage. One very important advantage is that access to data stored on a system’s edge is not limited during periods of peak traffic on backbone and regional distribution networks.

The Bottom Line

Locating data caches for Live Replay and related DVR tasks on your systems’ edges is now a well proven solution to many complex cloud DVR system issues. In conjunction with soft upgrades to add well-designed User Interface software, this approach can be retrofitted to existing systems utilizing the SD, micro SD or USB form factors or designed in with the appropriate high endurance e.MMC device.

REFERENCES

1) Time limits for echo tolerance in real-time — “Understanding the Echo ‘ Phenomenon’ Causes and Solutions”

2) Typical tolerance to service delays — “Response Times: The 3 Important Limits”
   https://www.nngroup.com/articles/response-times-3-important-limits/