BROADCAST THE BRIDGE
Connecting IT to Broadcast

Delivering Timing For Live Cloud Productions

Essential Guide

EG
ESSENTIAL GUIDES

Introduction

By Tony Orme, Editor at The Broadcast Bridge

IP is an enabling technology, not just another method of transporting media signals. Consequently, it is giving broadcasters the opportunity to reconsider how we build television workflows and infrastructures.

One of the greatest strengths of television is the concept of backwards compatibility. This has allowed new generations of television technology to be introduced without compromising the status quo. The introduction of color was probably the greatest example of this when chrominance signals were modulated onto the luminance signal. This not only provided a single transmission format that was backwards compatible with the existing monochrome televisions, but also facilitated a whole new and exciting format for those wanting to watch in color.

From color we moved to digital broadcasting which facilitated the new widescreen formats with HD closely following it. Both systems provided backwards compatibility to the existing formats allowing viewers who wanted to buy into the new technologies the freedom to do so without affecting the existing viewers. And this has continued through WCG and HDR, with each new format maintaining backwards compatibility.

Maintaining backwards compatibility may well allow viewers to keep their television sets for tens of years, but it does so at a price. And as we transition to IP, the price we pay for maintaining backwards compatibility is having to maintain nanosecond timing, a consequence of providing color sub carrier for color broadcasting.

Television is an illusion, there are no moving pictures in television, just a series of still images that are played very quickly to give the illusion of motion. And it is this illusion that we must maintain to convince the human visual system that motion exists. But more importantly, fluidity of motion must be maintained as humans are exceptionally good at detecting timing anomalies, especially regarding motion.

IP, by its very nature is a packet switched network that does not necessarily respect packet order or timing. Packet delivery, especially in unmanaged networks such as the public internet are bursty and this is often exasperated by switcher and router buffers. Therefore, we cannot rely on the temporal separation of packets to deliver nanosecond timing.

As more broadcasters flex the power of IP, they are realizing that remote operation is becoming easier and more cost effective. But to truly empower IP and remote working we must make better use of the internet, unmanaged networks, cloud and datacenter computing. And to achieve this we need to take a different approach to timing. As any seasoned broadcast engineer or technologist will tell us, we must work with what we have and not what we want.

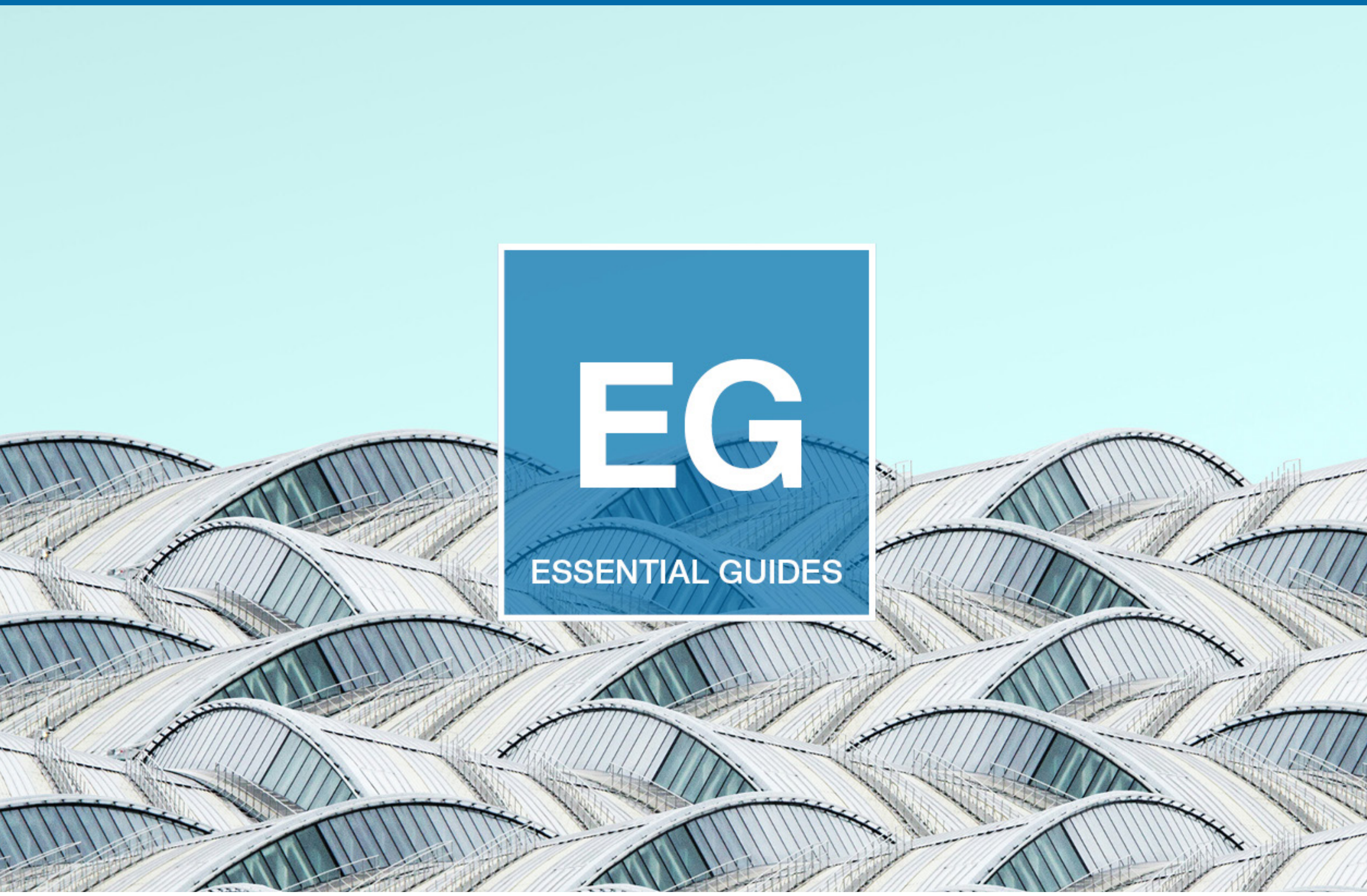
The history of television formats has been largely driven by the available technology of the time often resulting in broadcasters working at the cutting edge of technology. But as technology has evolved, we now have the opportunity to take a step back and look at broadcasting from the viewers point of view.

Tony Orme.

To improve the immersive experience, does the viewer still need nanosecond timing to maintain backwards compatibility with color subcarrier timing? If we can say “no” to this, then IP, cloud, and datacenter solutions are going to present us with untold freedom and opportunity.

Tony Orme
Editor, The Broadcast Bridge

Delivering Timing For Live Cloud Productions



By Tony Orme, Editor at The Broadcast Bridge

Video and audio signals represent synchronous sampled systems that demand high timing accuracy from their distribution and processing infrastructures. Although this has caused many challenges for broadcasters working in traditional hardware systems, the challenges are magnified exponentially when we process video, audio and metadata in software.

Software isn't necessarily the root of the problem when we consider timing in computer systems. Instead, the infrastructure the software runs on, such as the operating system, virtualized hardware, and networks, all contribute to creating indeterminate latency and timing errors.

Broadcasting is still saddled with the historic decisions made for the adoption of color broadcast, especially regarding nanosecond timing tolerances. Synchronizing pulses are no longer needed but broadcast engineers the world over still have a nanosecond mentality that is completely outdated and no longer needed.

Overcoming the limitations of unpredictable timing planes needs a new way of thinking about latency in broadcast infrastructure.

Viewer Expectations

Every so often in the development of an industry we have the opportunity to look at the technology and revisit how we deliver our products. In terms of television, our products are highly immersive programs that entertain, educate, and inform our audiences. Therefore, the technology should serve to deliver the product, not the other way around.

From the viewers perspective, they don't really care how a television program is constructed or delivered to them. We don't really care how a letter is mailed across our respective countries, and the same is true for program delivery. It's fair to say that viewers have expectations, which manifest themselves in terms of constraints of the system, but the actual technical detail of how a letter posted in London reaches a home in Sydney is largely irrelevant.

Our ever-demanding expectations of the postal service very quickly influences the type of technology that is employed. For example, many moons ago it would have been acceptable to send a letter by ship which would take six weeks to traverse the world, now we expect our letter to travel by air mail and be delivered in a few days. And then when we look exactly at what a letter is, we realize it's just a form of communication, in other words it's information which can be represented as an email, which only takes a few seconds to arrive.

Although delivery times for television have always been in the order of seconds, the analogy to mail delivery is similar, but instead of the time-to-delivery changing on the part of the mail user, now we have viewers who expect to watch what they want, where they want, and how they want, with a constant pressure on reducing costs to them. It's this expectation that has now placed a constraint on the broadcast system, which in turn has demanded we provide new technology.

Viewers are making these demands and we need to find a new way of delivering for them. And the great news is that we have a solution, it's called IP with its flexibility, scalability, and resilience. However, the devil is always in the detail.

Nanosecond Timing

Before understanding why, we need to completely rethink timing, but in doing so it's worth reviewing where we are and how we got here.

There are two points to remember; there are no moving pictures in television, just a series of still images played very quickly to give the illusion of motion, and we are still using the same timing constraints that were designed in the 1930s and 1960s to overcome the limitations of the technology of the time.

Both electronic cameras and televisions of the 1930s used vacuum tube technology and they were truly scary as EHTs (Extra High Tension) of more than 14kV and high current circuits were the norm. Mainly, this was due to the need to direct and project electron beams. Sync pulses were needed to shift the beam left, right, up and down the screen (or camera sensor) resulting in massive currents energizing scanning coils.

Another name for these is an inductor, leading to very long sync pulses to not only move the beam back to the start of the line, but also keep it sufficiently temporally long so that the change of current didn't destroy the driver circuits when generating the back EMF (Electromotive Force).

Furthermore, when color started to appear in the mid 1960s the concept of color subcarrier was introduced to maintain backwards compatibility with existing television sets. This cemented the need for nanosecond timing tolerances to allow the QAM (Quadrature Amplitude Modulation) demodulators to decode the color in the televisions along with demodulating the audio. But our reliance on analog television has been reducing as digital television becomes mainstream.

All these systems were needed in the 1930s, right up to about ten to fifteen years ago, when viewer expectations were relatively modest compared to today. But as digital transmissions progressed and flatscreen televisions and mobile devices started to appear, viewer demands increased exponentially to the heights of where we are now, meaning that we must constantly innovate new solutions to deliver for our viewers.

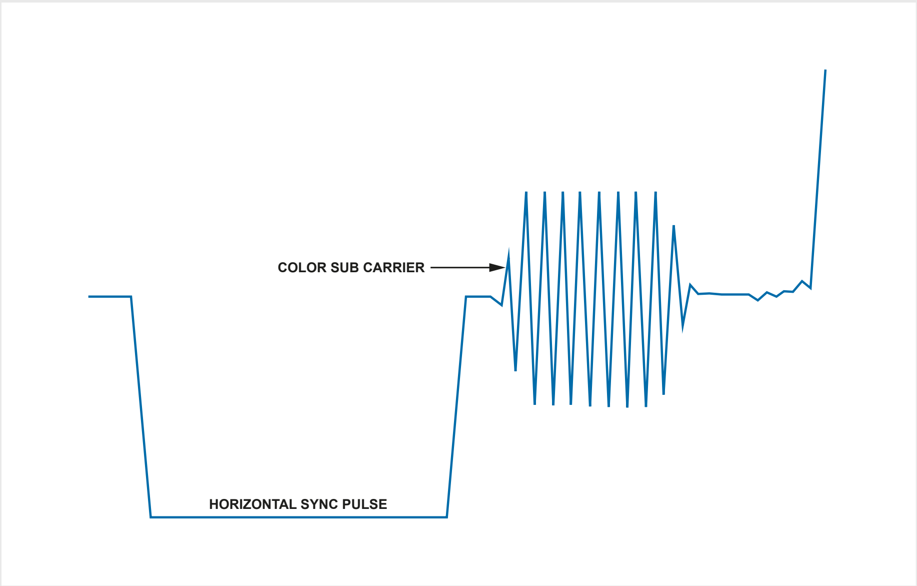


Figure 1 – Horizontal line pulses were devised in the 1930s to synchronize electron beam scanning cameras and monitors, and color subcarriers were created in the 1960s to provide backwards compatibility for black-and-white TVs with the introduction of color. Neither have been needed for at least twenty years, and with the adoption of IP, we now have the opportunity for change to allow us to deliver a better immersive viewing experience.

Remembering The Viewer

The HVS (Human Visual System) is a system that is greater than just our light transducers, otherwise known as our eyes. It encapsulates a whole behavioral psychology that has not only influenced television design but driven it. The eyes provide visual data and prompts for the brain, and the psychology of the brain provides our internal representation of the image. For example, we must simulate fluidity of motion otherwise the HVS detects motion anomalies that can be interpreted as a predatory attack due to flicker. Even when the viewer is sat in the comfort of their own home and know they are safe, disturbances in the fluidity of motion can lead to stress, and that's before we even start talking about the psychological effects of disturbances in sound.

From the perspective of our viewers, we must make sure the images are smooth and flicker free, as this adds to the immersive experience.

All this considered, we don't need color subcarriers anymore. All we need is a reference to pixel 0 of the image and an idea of the frame rate we're using. If we treat the image as a matrix consisting of 1920 x 1080 pixels (for HD), then it's easy to see why we only need to reference the first pixel as every other pixel forms part of that matrix and can be easily determined.

Maintaining Fluidity Of Motion

To keep images fluid for our viewers they must be displayed with a consistent and predictable time-base. If its erratic, speeds up, or slows down, then this will trigger the ancient structures in our brain which form the HVS resulting in stress for the viewer, manifesting as a lack of immersive experience. But it's important to remember that the image frames do not necessarily have to be transferred with a constant time base, just displayed that way.

This is a very important step in the evolution of broadcast television as we're now moving away from the timing constraints imposed by the technology of the 1930s and 1960s, and hence the constraints on the viewing experience. We no longer need to worry about scanning coils and back EMF, but we do need to be concerned with maintaining the viewer's immersive experience.

A)

Oscillator tolerance equation is:

$$\frac{f_a - f_b}{f_b} = ppm \text{ (parts per million)}$$

Assume an HD pixel clock at 74.25MHz with a 150ppm tolerance, then the HD pixel clock frequency f_a can increase to:

$$f_a = (ppm * f_b) + f_b$$
$$= (150 * 10^{-6}) * (74.25 * 10^6) + (74.25 * 10^6)$$
$$= 74.26113765 \text{ Mhz (thus, leading to a relative frame rate increase)}$$

B)

$$\text{frame rate} = \frac{f_a}{Pixel_{horiz} * Pixel_{vert}}$$

$$\text{frame rate} = \frac{74.26113765 * 10^6}{2640 * 1125}$$

$$\text{frame rate} = 25.00375Hz$$

Therefore, a frame rate increase of 0.00375Hz

$$\text{Time difference} = \frac{1}{0.00375} = 226 \text{ seconds (4.4 minutes)}$$

Figure 2 – A) shows how a 74.25Hz HD pixel clock with a 150ppm tolerance increases in frequency. B) The frequency change relative to a 74.25Hz HD pixel reference clock leads to either too many frames being generated, or too few. If the clock runs fast, then video frames will need to be dropped every 2.2 minutes (in the worst case), and if it runs slow then video frames will need to be duplicated every 2.2 minutes (in the worst case). The 4.4 minutes represents only one clock running fast at +150ppm, but the clock it's running relative to could be running slower at -150ppm, hence the 4.4 minutes is divided by 2 to give the worst case of dropping or duplicating video frames approximately every 2 minutes.

And this gives us the freedom to think about timing differently. Instead of thinking in terms of what the technology can provide for the viewer, we need a mind-shift and ask the question "what does the viewer want and how do we deliver?"

When we say, "constant frame rate", what do we really mean? In practical terms it is impossible to reach exactly 50Hz or 60Hz, but we can generate frame rates at these frequencies with a certain tolerance, hence the reason we have sync pulse generators and fly-wheel oscillators that lock to the reference signals. At this point, it's easy to disappear down a rabbit hole and start making our reference generator more and more accurate, thus decreasing the timing variance to achieve nano-second tolerance, so that the oscillator becomes incredibly accurate.

One reason for our strict timing is to synchronously switch between video and audio sources by making them frame synchronous. And again, we should ask what exactly do we mean by "frame synchronous?". In the NTSC and PAL days we would tweak the SCH-phase to adjust the line timing into the production switcher to make the video sources line and frame accurate, and this was necessary as the alternative was to use frame synchronizers and they were hugely expensive. When digital switchers matured, they had line buffers built into every input so that the timing tolerance only needed to be plus or minus a few lines. Nobody has tweaked an SCH-phase in an SDI broadcast center for about ten years, but we still talk about nano-second timing.

Video Frame Referencing

As an alternative, we instead think in terms of frame referencing, then we have a system that is much easier to work with, is much more flexible, and delivers many more infrastructure options.

This can be achieved by timestamping each frame, but instead of using a frame synchronizer to line each video input to a common sync-reference, we instead change the offset in the timestamp. Then we find that we have a system that is no longer reliant on clock synchronous transport streams such as SDI and can work in IP COTS infrastructures as well as public clouds.

One timing reference could be derived by synthesizing the optimal timing point. And assuming all the other video sources had a similar frame rate, then they would be displaced temporally on the timeline relative to the reference input. Each video input timestamp would be normalized to match the nominated reference video and adjusted so that each frame aligns in time. It's clear we cannot move a video frame in time, but we can send it to a buffer to be delayed to temporally match the nominated video source reference.

The playout-engine will have visibility of the contents of each buffer and will be able to read out the appropriate frame at the designated timestamp. And by choosing an input that has the "oldest" timestamp, we can then use the buffer to hold back the other input frames. In effect we've created a one frame buffer, but instead of moving the video frame through the buffer's memory, which is incredibly resource hungry and therefore inefficient, we change the pointer, or timestamp so the playout-engine knows where to retrieve the video frame from. This results in a system that is highly efficient as we're reducing the amount of data we are moving around the memory.

The challenge is that the streams of video frames are not locked, they are asynchronous relative to each other. If we take one to be our reference, the other streams will create more or less frames than the reference in a time period. However, this system assumes the frame rates are similar, which is a reasonable assumption in broadcast television. They're not the same, as we would find with an SPG (Sync Pulse Generator) locked infrastructure, but they are close enough. Figure 2 shows that a standard off-the-shelf 74.25MHz oscillator with an accuracy of 150ppm used as an HD clock source in a camera, exhibits a frame drop or frame duplication once every two minutes.

Would a viewer at home see this? Well, we know they don't because this is how a frame synchronizer works when used to synchronize an outside broadcast video feed for a studio. The adding and dropping of frames are processed differently for a contribution feed in the studio compared to the received signal at the consumer. The CODECs employed are wonderful at smoothing frame anomalies so viewers don't see them.

Key to understanding this is remembering what we are trying to synchronize and why. In the IP world, we no longer have to frequency and phase lock color subcarriers or SDI transport stream clocks. Instead, we just need to maintain fluidity of motion, which can be achieved at the frame layer where tolerances are much more forgiving.

Operational Latency

Another area where we can think differently about time is human response times. When switching between video sources there will be some delay. In SDI broadcast facilities we didn't give this much thought as all the signals and control equipment were relatively close and the propagation times of signal paths were very low. However, as we move to IP and internet operation, we cannot take these latencies for granted and must take a closer look at operational controls.

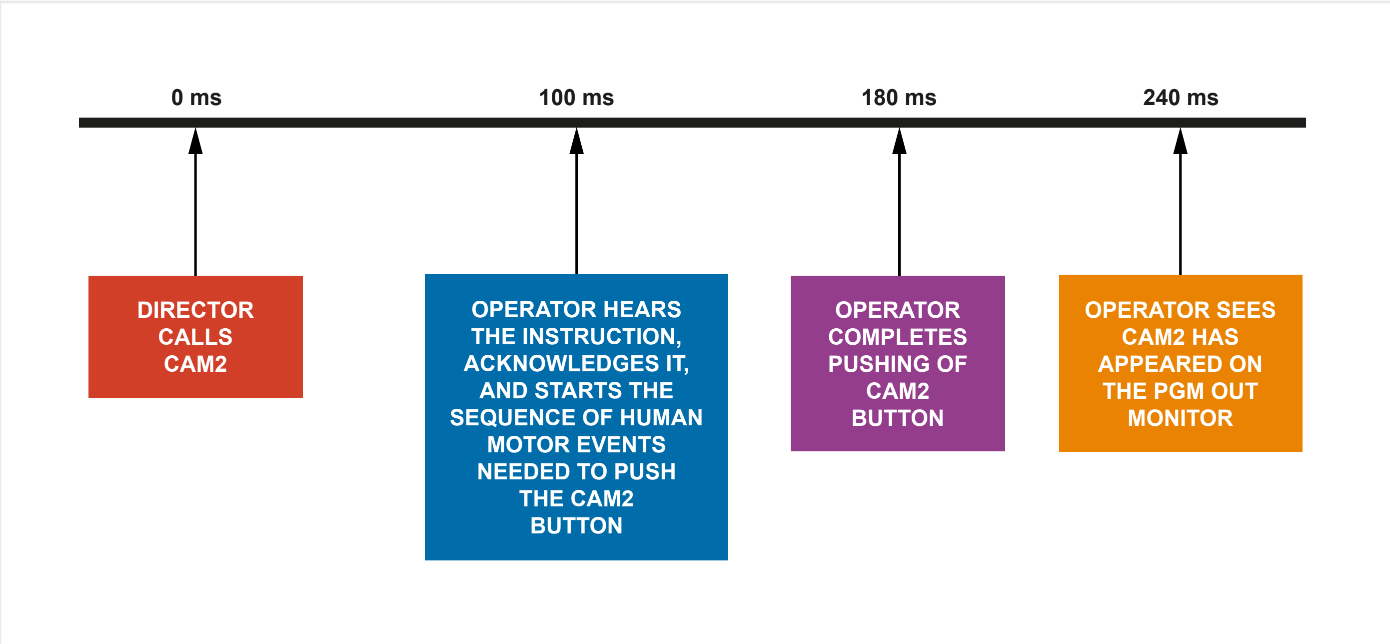


Figure 3 – Human response times are much longer than we may think. This provides the opportunity to integrate network and cloud processing latency into the workflow without any noticeable effects.

There is a tendency to over-use the word “instantaneous” in relation to switching response times, especially when controlling equipment such as production switchers. We like to see an “instantaneous” response time when switching between video sources, but the switching response time has never been instant, there’s always been some delay between switching an input on the production switcher and seeing the change on the program monitor.

Research has demonstrated that the average human takes about 240ms from the triggering of an event to recognizing the response. For example, as seen in Figure 3, if the director calls for the operator to switch from CAM1 to CAM2 on the program bus, the operator, on average, will not recognize the change in video output on the program monitor for about 240ms, or 8 frames of 30fps video.

The profound impact of considering human factors is to realize that it is in fact human factors which can lead to the best time base management of audio and video streams. Because these streams are buffered, and include time stamps, which can be as simple as the RTP stamp for top of frame in SMPTE ST-2110, or as powerful as PTP stamps, used in the same RTP example, it is possible to manage overall latency for a worldwide distributed system.

This opens a whole load of new possibilities for remote controlling video processing equipment. If we now reframe our definition of “instantaneous” not in terms of nano-second timing, but in terms of video frames, we can see that the phrase “instantaneous” now means 8 frames of 30fps video. Consequently, operational controls over the internet can be achieved in many cases. In other words, the operational latency must take into consideration the actual expectation of the users, in this case, the production switcher operator.

Conclusion

Timing in live production environments has always been something that we’ve strived to improve. Nano-second timing and near-zero latencies have been assumed to be fundamental requirements but as we’ve seen, these are based on historical technological constraints and assumed folklore. When we question our assumptions then it becomes clear that there are simpler and more effective ways of working to achieve scalability, flexibility, and resilience, especially in the world of IP.

The Sponsors Perspective

Unchaining Time

By Chuck Meyer and Chris Merrill

What is real time? While that question doesn't normally come up at the dinner table, asking it of a group of broadcast engineers can draw out all kinds of responses, from philosophical debates around global atomic clocks to technical dissertations on lines, frames, and permissible nanoseconds of processing delay.

One of the reasons there are lots of opinions on the topic is because time is a human construct that we use for sequencing events. Real time describes a human sense of time that seems immediate.

The perception of real time – what is happening in a specific moment – is heavily influenced by what is happening in a person's environment when they perceive it. Therefore, the definition of what is real time can vary by individual.

Before we start getting all metaphysical, let's narrow the discussion. In live media production when we talk about working in real time what we are really asking are two separate questions.

1. Is there is a noticeable difference between when I perceive something happening and when I can act on it? This is relative latency. A system that feels live to the operator must have a response time of about 240 milliseconds from the time the operator sees the cue to seeing the result of the action they have taken.

2. Using a 24 hr clock, how many seconds does it take to sequence the different processing steps taken on a frame of video before it is pushed to the viewing audience? This is absolute latency. The expectation for absolute latency varies widely by producer but usually is less than 30 seconds.

The reason to break this into two separate questions is because if all the processing steps involving relative latency can be properly sequenced within the expected absolute latency, it doesn't matter how many there are or when they occur. The system operators will take their actions in what feels like real time and the audience will have a live viewing experience.

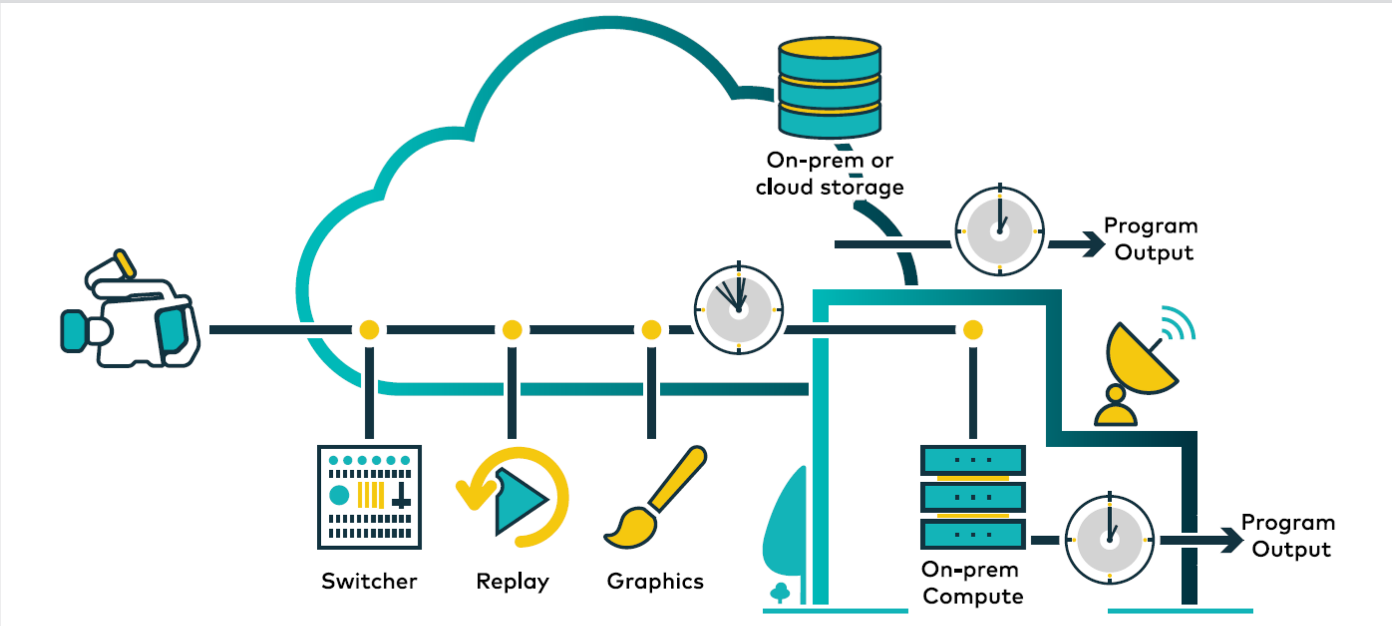
To see how this works, let's look at AMPP, Grass Valley's Agile Media Processing Platform. In AMPP, every video frame is timestamped as it enters the system. Because transport times vary as frames speed across networks to different members of the production team, AMPP also tracks the local time of each operator. This allows creative decisions made by the operator and their associated processing time to be tracked relative to the operator's time. The result of the operator's work is time stamped with whatever offset time is best to synchronize the work across the production chain.

With AMPP managing these timing offsets, the operator experiences the phase-aligned environment they are used to. The order and local timing of the decisions are maintained. When all operator actions are sequenced, the total environment is time-shifted relative to the source and thus maintains the program's continuity.

Following this design strategy, any live production task can be carried out in what feels like real time and assembled in a linear fashion to create programming that exceeds audience expectations. Even with complicated production tasks, total execution time is a few seconds. Compare this with today's traditional live broadcasts which, in the best of circumstances, still take as much as 50 seconds to get final emission delivery to the home.

Unchaining individual operator workstations from external time is possible because AMPP operates faster than real time using technologies that did not exist when traditional frames per second timing was implemented. Frame syncs that were once used to introduce a few frames of delay are replaced by memory buffers which can hold the frames until they are needed for the sequence.

AMPPs internal frame management allows unique offsets for each operator by adjusting the buffer depth to match the timing offset required for each essence or AMPP can force groups of operators to be synchronized if that timing is critical to their workflow. In either case the perception of the operator is that the system is responding to them in real time.



New technology can align contributions from multiple contributors.

Dennis Breckenridge, CEO of Elevate Broadcast Pte Ltd described their experience with AMPP in this way:

"With our virtual product we went whole hog. We had no backup plan. We counted on AMPP fully to work and we pushed the boundaries.

"We had contribution from many different countries: Australia, Singapore, the Philippines, Indonesia, and Thailand. Our producer was in Singapore. The director and TD with the switcher were side by side in Sydney, Australia. The main cameras were all in green screen studios with virtual sets but we also had live Zoom feeds and other complications.

"We told the production team: 'You can't come to Singapore because of the pandemic. You can stay there and we're still gonna make everything that you're used to: Karrera panel, multiviews, comms… All these type of things we're gonna make magically work for you and you'll produce a major broadcast for Asia!' It took a little time to build their confidence and acceptance of that possibility.

"Once all the comms and everything came together, the concerns from the production team went away. We managed all the delays through the system. Once that happened, they forgot about the technology and they just moved on with their production. That was the end of it. They felt like they were just in two different control spaces within the same facility. They didn't think about the fact that they were on different continents."

AMPP manages both relative and absolute latency in a way that makes the difference invisible to the operator and audience, erasing the barriers that were previously very apparent in remote production.

Chuck Meyer.

Chris Merrill.