Time-Compensated Remote Production Over IP

By Ed Calverley

Abstract

Much has been said over the past few years about the benefits of moving to use more Internet Protocol (IP) in broadcast, most of which has focused on simply replacing the existing serial digital interface (SDI) connections with IP ones. This paper will look at where the use of IP can enable innovative ways of working that would not be possible or practical without IP. This paper will pay specific attention to remote production as this is an area where latency cannot be avoided. Finding methods to make use of the latency can lead to more flexible methods of production, which could drastically change the costs models for live production of outside events.

Keywords

BBC Sport, distributed production, Euro 2016, Iphrame, IP networking, IP production, live production, outside broadcast, remote architectures, remote controlled, remote production, remote sources, suitcase TV, switching, time-compensated, TimeLock, vision mixing

What Is "Time" Anyway

The terms *Time* and *Timing* mean many things to many people in broadcast but, whatever the interpretation, understanding exactly where a frame of video or sample of audio belongs is what makes television work. For many, time can simply mean *time of day*, for others timing refers to the measure of frequency and phase. Systems generating video frames, sampling analog signals, or handling multiple signals together all rely on having some sort of reference signal to enable them to derive accurate frequency and phase alignment to ensure their processing occurs at a predictable and stable time over lengthy periods of operation.

Since the introduction of high-definition broadcasting, it is becoming common for systems to be designed to handle a mix of frame rates (e.g., 1080i/25 and 1080p/50). With increased sharing of media on the global market as well as online, more complicated mixes of frame rates can be encountered too. It is typical in systems designed to handle a variety of frame rates that time is measured as an absolute value with at least millisecond accuracy.

SMPTE ST2059 defines how the IEEE's Precision Time Protocol (PTP) should be used in broadcast systems. PTP time addressing provides a mechanism for identifying time down to nanosecond granularity relative to the defined epoch of 1 January 1970. The standard defines how PTP values equate to timecode labels commonly used in broadcasting and the expected video signal phase to be calculated for all standard video formats. Essentially, the use of PTP clocks on an IP network replaces the need for both time-of-day timecode distribution (e.g., linear time code) and other reference signals (e.g., BlackBurst).

This paper will look at where the use of IP can enable innovative ways of working that would not be possible or practical without IP.

"Realtime" Processing

The early days of television broadcasting certainly relied on "realtime" processing. The scanning electron beam in the tube of the cathode ray tube (CRT) television set directly followed that of the scanning in the tube inside the camera. Signal path switching and vision mixers had to be carefully designed to maintain a consistent scanning raster. To achieve this, all devices would be referenced to operate at the same processing frequency and have their processing phase carefully adjusted to ensure signals arrived at any switching point at a very carefully controlled time.

While signal flows through digital television studios, master control, and transmission systems are considered to be realtime, in reality there are many places where signals are artificially delayed ensuring the "realtime" behavior is correctly maintained.

As software-based processing and commodity IT-hardware becomes more common in the broadcast chain, these small delays are starting to increase; primarily since most software-based systems process video one frame at a time and typically have frame-based input and output buffering resulting in the total throughput latency being measured in frames rather than in lines.

Digital Object Identifier 10.5594/JMI.2018.2813199
Date of publication: 1 June 2018

Production environments that have implemented virtual sets or augmented reality graphics have already had to learn to work-around significant delays in video feeds from a few frames to a couple of seconds. Choices must be made to delay the audio and other video feeds so that everything remains in-sync in the gallery or whether to delay signals downstream and accept a lack of lip-sync in the gallery. A crucial factor in this decision is the use of open talkback in a gallery where audio spill from the production may get back to a presenter's earpiece—with just the right (or wrong!) delay, some presenters may be rendered incapable of speaking (as shown in the overly simplified **Fig. 1**).

The move to IP doesn't directly imply an increase in processing latency; where native uncompressed IP interfaces and nonblocking network switches are used, the latencies are comparable to operation with SDI. However, the use of IP does increase the chances of more frame-based processing (e.g., software-based systems) being introduced which may be more likely to increase the overall system latency.

So, while live production will continue to be considered a "realtime" process, the processing latencies through the various signal paths can never be removed and to ignore them will eventually lead to issues. A better option may be to embrace the latency and to use it as an advantage to enable innovative ways of working as this paper will outline for remote production.

Why Do We Need Remote Production?

Coverage of live events that are held away from a production center can be very expensive due to a range of factors. In addition to technical facilities to capture and mix sound and pictures, there are typically teams of people ranging from 1 to 100+ depending on the size of the event. The costs for travel and transport, accommodation, and subsistence can make it uneconomical to cover some events.

Coverage of live events is a great way to attract and retain viewers, whether broadcast live on a linear TV channel, streamed live online, or packaged for access through on-demand platforms. As viewing habits are changing and people are consuming content in new ways, viewers expect a wider choice of content and broadcasters can struggle to provide this with budgets being continuously squeezed.

Sports broadcasters may have paid significant amounts to acquire rights to cover certain high-profile events which are often packaged with rights for a range of events, many of which will never get televised as there is no commercial justification to cover the expense of doing so.

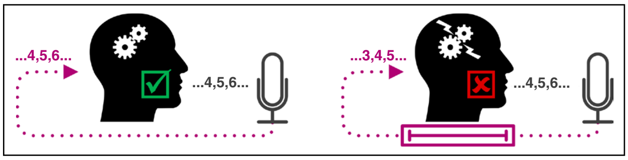

FIGURE 1. Audio spill through open talkback can be problematic if delayed.

The main driver for remote production is to reduce costs, the priority being a reduction in people onsite as the facilities costs may be small in comparison.

If the correct architectures are chosen, the cost savings may be significant enough to make even low-profile events economical to cover. Compromises on operational flexibility may be necessary with some architectures, but these must be weighed up against the cost savings. Some architectures may not be achievable with traditional broadcast hardware, but in an IP-based world more options are becoming available and new hybrid solutions will likely become common very quickly.

Remote Production Architectures

The architectures discussed in this paper all focus on the relocation of operational/production teams. In most cases (except for fixed installations), engineering staff would still be required onsite to set up and manage equipment.

Remote camera operation and racking is not discussed in this paper but it is also an area where IP technologies can provide benefit. Arguably, both operations are significantly more challenging to compensate for any latencies present in monitoring feeds. However, depending on the type of coverage, the use of IP-controlled pan-tilt-zoom (PTZ) cameras or mounts with recall of preset positions may further reduce onsite effort.

With all remote production architectures, there is an obvious risk factor which may force a broadcaster to continue with a traditional outside broadcast operation; when produced locally (onsite), the final output (and any ISO feeds) can be captured locally, meaning that even in the event of a major link failure, the event coverage is safe (i.e., can still be used for playback later). With remote production where there is no backup capture or mixing onsite, any interruption in the link may result in complete loss of coverage, jeopardizing revenues, and is likely to have a negative impact on a broadcaster's reputation.

Architecture 1: Production With Remote Sources

The simplest form of remote production is arguably not remote production. Major sporting events at fixed venues have already justified the investment in dedicated fiber links specifically for use by broadcasters to return multiple feeds to production centers either uncompressed or using very light compression.

Rather than sending an expensive production team to an event location and having full video and audio production facilities in a mobile unit, many broadcasters are experimenting with returning all sources allowing fixed-facilities at a production center to be used (**Fig. 2**). Importantly, this means staff get to go home at

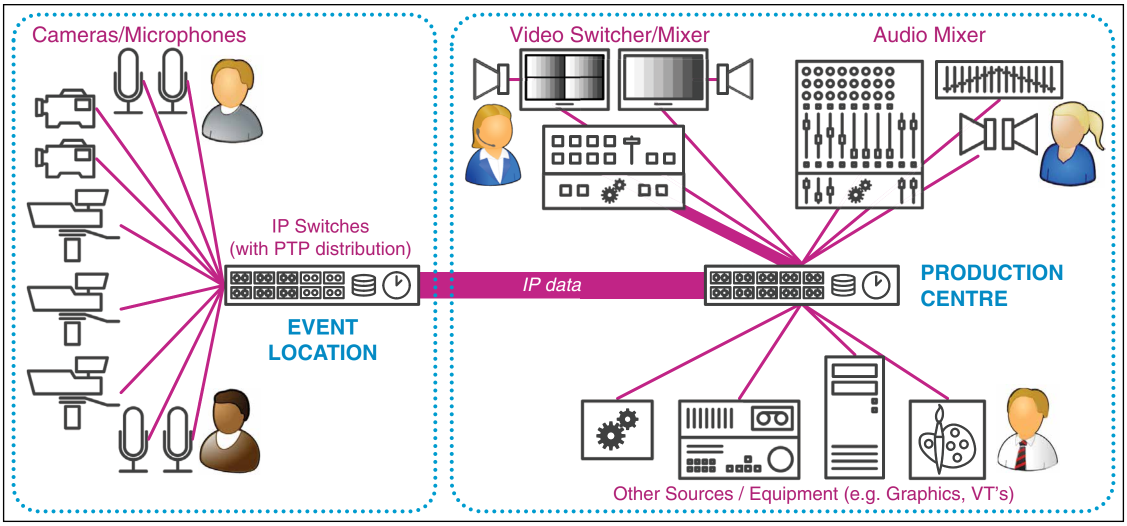

FIGURE 2. IP-based production with remote sources.

night and potentially can work on coverage of multiple events on the same day.

This architecture is the simplest and perhaps allows even more flexibility than traditional outside broadcast productions due to the potential increase in facilities available in a production center.

An increase in the number of sources will result in a linear increase in the link bandwidth required. This architecture is therefore only practical where high-bandwidth links are readily available and production budgets can cover the costs.

It is common for some outside broadcasts to have an active backup link or some emergency way to get a feedback if the primary route fails. With the higher link bandwidth required for this style of remote production, the costs of a backup link can become significant.

For high-profile events with large production teams, the cost savings gained by not having people onsite can make this very attractive even though the costs of links may be high.

Architecture 2: Remote Controlled Production

When budgets are squeezed, one option for keeping operational staff at the production center without requiring significant link bandwidth is to keep the vision mixer processing onsite. This model can work over much more limited network links as typically only two video feeds need to be carried back.

The remote vision operator would use an IP-connected control panel with the main mixer processing unit being onsite. They would monitor the following video feeds:

- source multiviewer
- mixer program output

Even with only light compression, both these feeds would suffer some amount of delay. **Figure 3** shows a best-case example where sources are delayed by four frames (this could be more, depending on how the source multiviewer feed is generated or depending on the compression and carriage mechanism used).

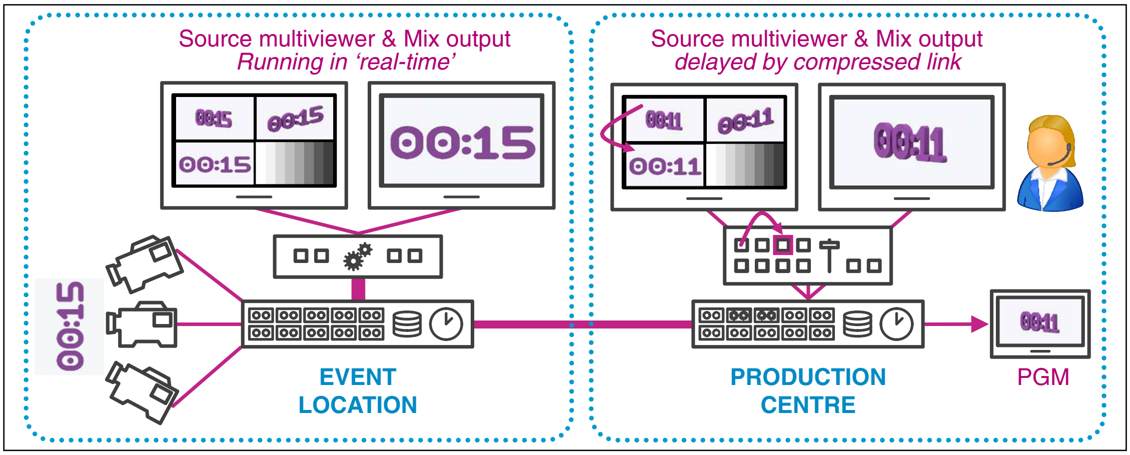

FIGURE 3. Remotely controlled vision mixing.

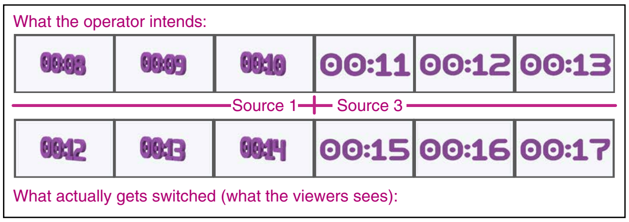

FIGURE 4. The impact of monitoring latency on remotely controlled vision mixing.

Figure 3 highlights that while the equipment at the event location is operating in "realtime," the operator's view will be showing frames in the past (e.g., operator sees frame 00:11 while onsite is processing frame 00:15). Assuming button presses on the control panel are relayed back to site with negligible delay, the result of any user actions would not be seen on the operator's program output monitor until at least four frames later.

As shown in **Fig. 4**, due to the monitoring delay what the viewer ends up seeing is not the same as the vision operator intended. For fast-moving sports where every frame matters, this inaccuracy in switching between sources could significantly impact the quality of the coverage making this model a significant compromise to accept.

Architecture 3: Time-Compensated Remote Controlled Production

The previous two architectures can be achieved with commonly available broadcast hardware and fundamentally aim to operate in a "realtime" way with all efforts placed on minimizing any latencies. New solutions can make use of timestamps common in the IP-based protocols giving more control over the time at which signals are processed allowing architectures that use the latencies as an advantage.

In addition to carrying uncompressed signals with PTP-based timestamps as with SMPTE ST 2110, a similar timestamping technique can be applied to lower resolution proxy versions of feeds. The resolution and compression used on these proxies can be adjusted to suit the link bandwidth available to return them to where an operator is located.

At the control location (typically the production centre) operational staff can view these proxies in sync with each other thanks to their embedded timestamps. This method for synchronizing and controlling the display time offset gives more flexibility for monitoring compared to a precompiled multiviewer feed and allows the operator to be located anywhere there is suitable IP connectivity.

By ensuring the systems at the event location and remote control location are both locked to accurate PTP clocks and by using a small amount of buffering, it is possible for the operator's view to be considered "as-live" with a defined fixed offset that is unaffected by any jitter on the network.

The time difference between sources being captured onsite to them being displayed at the operator's location is the latency measure that matters, as it can impact the ability to give verbal direction to camera operators and/or presenters onsite. Typically, this latency must remain under 1 sec if responsive direction is required. In **Fig. 5**, the operator's monitoring is shown running with a five-frame offset (i.e., the pictures the operator sees were captured five frames ago).

To compensate for the monitoring latency and to ensure accurate vision switching can be performed, the mixing process must operate at a time offset larger than the overall round-trip latency. **Figure 5** shows the full-resolution mixer running at a delay of 10 frames—this is the product of the five frames monitoring latency plus the time taken for control messages to be returned to the event site, and appropriate buffering and source signal processing time to allow frame accurate processing.

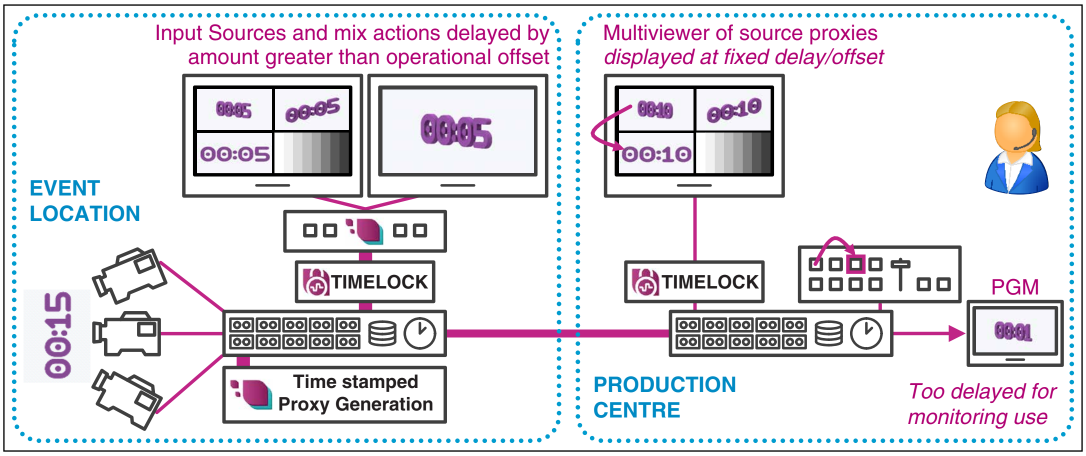

FIGURE 5. Remotely controlled vision mixing with a fixed processing offset of five frames.

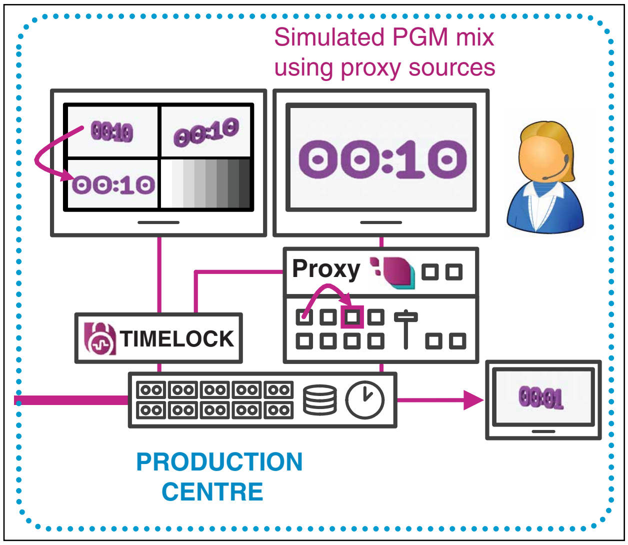

FIGURE 6. Simulated mix process using proxy sources.

The boxes labeled "TimeLock" in **Figs. 5–7** are Suitcase TV's commercial name for this combined software process, which synchronizes multiple feeds and aligns/delays them to an exact offset (referenced to PTP) in one software process. This architecture can be visualized in hardware as a simple delay being applied to signals feeding into a mixer/switcher (assuming all signals are carried in perfect sync relative to each other).

This architecture ensures that switching is done on the correct frame intended by the vision operator. With the additional processing delays used, it is not practical for the operator to rely on monitoring the output feed from the mixer at the event location (e.g., returned over an IP link) as they would have to wait an unacceptable time to see the effect of their actions.

To provide a better experience for the operator, the timestamped proxy feeds can be used to perform a "realtime" mix locally at the production center, providing a simulation of what the onsite mixer will be doing slightly later (**Fig. 6**).

A similar process using the proxy sources could also be used to generate a preview output to provide the full program/preset behavior expected by vision operators (removing the need to return a preview feed from the onsite switcher/mixer).

While this architecture is technically possible with today's hardware, it relys on careful adjustment of the timings and possibly cannot cope with jitter on the link so would be hard to guarantee frame accuracy.

To achieving this architecture in traditional broadcast hardware, the proxy monitoring and simulation mix setup may be an existing studio installation where the proxies are HDSDI signals carried over heavily compressed links. Actions performed on this proxy switcher/mixer have to be sent over an IP-link and repeated on the remotely located (uncompressed) mixer/switcher. To compensate for the delays in the generation and transport of the proxy feeds, the uncompressed sources need to be carefully delayed to match the timing of the commands being received.

Architecture 4: Distributed Time-Compensated Production

The previous architectures assume all sources are originated at the same location. If operational staff are remotely controlling a production, it is likely that some sources may be originated at their location (e.g., third-party graphics and video clip playback); clearly it would not be practical to transport these feeds out to the event location to be fed into the vision mixer there.

The solution is simply an extension of the same "time-compensated" concept outlined in the previous architecture. Any sources originated at the production center would be timestamped against the same PTP reference and have proxy versions generated that can be fed to the operators' monitoring in the same way as remote sources (with the same offset).

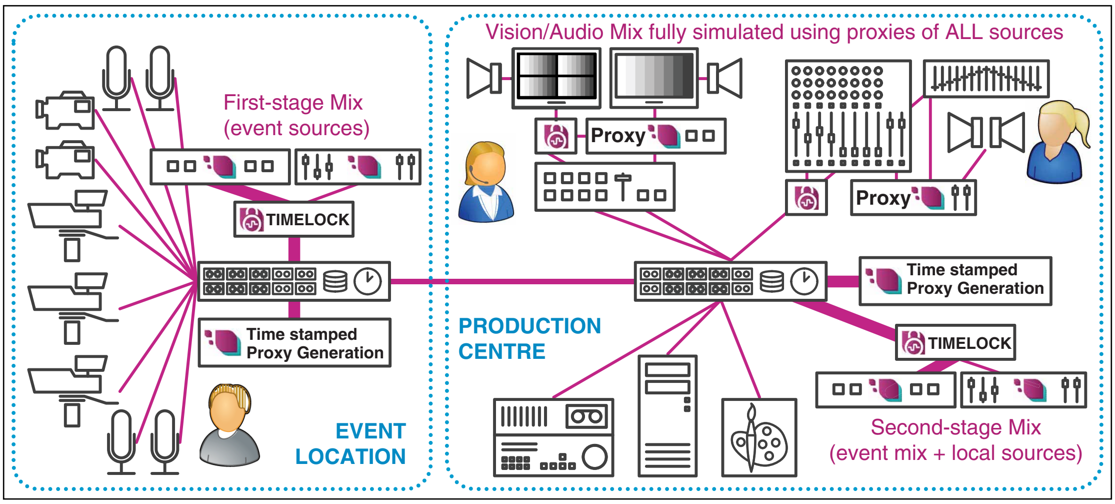

FIGURE 7. Distributed remote production (multistage remotely controlled mixing).

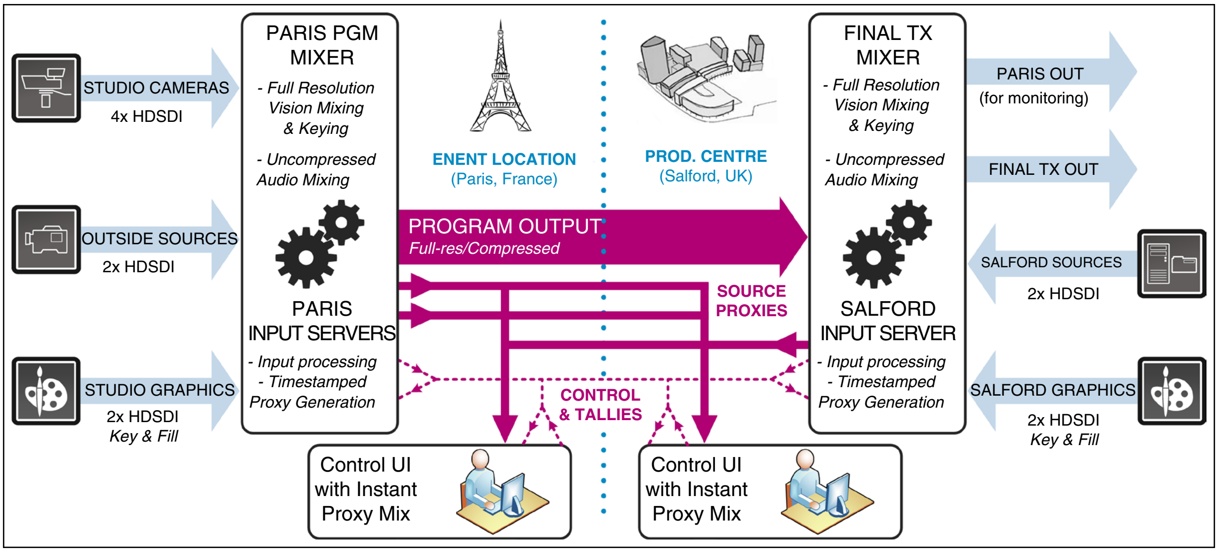

FIGURE 8. Distributed production at Euro 2016 (multistage remotely controlled mixing).

A second set of full-resolution mixing would then be performed to mix between the local sources and the feed from the event location (**Fig.** 7). This downstream mixer would run with a larger processing offset than the one at the event site to allow time for the event mixer program feed to be received.

Remote Production Trial at Euro 2016

In June 2016, Suitcase TV partnered with BBC Sport to perform a remote production trial during the Euro 2016 event in Paris. The trial implemented the distributed architecture described in **Fig.** 7 with specific signal architecture as detailed in **Fig. 8**.

Sources at the event location in Paris were mixed onsite using a software-based mix process and 10 GbE networking between processing machines. Compressed proxies for each source were carried over an IP network back to the U.K. alongside a single full-resolution feed carrying the program output of the onsite mixer.

At the production center, a second software-based vision mixing process was run that switched between the feed from Paris and locally originated sources. The sources from the production center were also sent back across the network to Paris so that operational positions at either location had the same view of all sources. Having all sources also enabled a simulated mix to be generated in "realtime" (i.e., following button presses) showing the operator the result of the action moments later in Paris and later still in Salford. The trial operated over a network with the bandwidth being as low as 50 Mbits/s.

The trial was an early proof of concept for this technology, so a lot was learned technically regarding the use of IP networks and the importance of having an accurate time reference at all locations.

Operational feedback from the trail highlighted the only specific latency measure that mattered between the sources being captured and the display at the control location—the practical limit being about 1 sec before direction of cameras and presenters over talkback would become problematic.

The greater offset/delay of the full-resolution mixing was not an operational issue as link delays are already a common issue feeding outside broadcast into master control/presentation. However, it was found that monitoring of the final mix output was deemed too distracting for operational staff who preferred to only see the simulated mix.

Conclusion

What is currently being referred to as remote production is simply a first step, making use of low-latency IP links to allow production to be moved from the truck parked outside an event location to a distant fixed installation (with the appropriate fiber-connectivity).

True remote production should not be underestimated as simply being "remote control" of equipment at a different location. To be useful in many productions, it requires methods for handling video and audio mixing of sources that originate at the production center as well as the event location—which is unachievable with a simple remote controlled architecture.

By compensating for latencies introduced by using software processing and IP links with limited bandwidth, distributed processing using multistage mixing can deliver viable architectures providing significant cost reductions. This provides opportunities for broadcasters to consider televising events which would be uneconomical with traditional outside broadcast methods.

Such innovations should lead to an overall increase in live content being made available for viewers with an ever-increasing appetite and expectation for variety.

About the Author

Ed Calverley started his broadcasting career joining the BBC in 1999 as a project manager, where he oversaw the digitization of a number of playout services. While at OmniBus (later acquired by Miranda, now Grass Valley), he was involved in the implementation of a number of large automation, newsroom, and MAM installations before becoming a key player in the delivery and product management for the new iTX platform, which brought about the channel-in-a-box revolution in playout. At OASYS, he continued to innovate in solutions for playout and helped move the company forward, significantly increasing turnover and ultimately leading to its acquisition by BroadStream. Now at Suitcase TV, Calverley continues to embrace new ways of using emerging technologies to empower the next revolution.

Presented at the SMPTE 2017 Annual Technical Conference & Exhibition, Hollywood, CA, 23-26 October 2017.