TIME-COMPENSATED REMOTE PRODUCTION OVER IP

Ed Calverley

Product Director, Suitcase TV, United Kingdom

ABSTRACT

Much has been said over the past few years about the benefits of moving to use more 'IP' in broadcast, most of which has focussed on simply replacing the existing 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. The paper will pay specific focus to remote production as this is an area where latency cannot be avoided but if it is embraced can lead to more flexible methods of production which could drastically change the costs models for live production of outside events.

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 analogue 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 HD 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 (ms) accuracy.

SMPTE ST-2059 defines how the IEE's Precision Time Protocol (PTP) should be used in broadcast systems. PTP time addressing provides a mechanism for identifying time down to nanosecond granularity with a time range of 136 years (2^{32} seconds). The standard defines how absolute PTP values equate to timecode labels commonly used in broadcasting and allows the expected video signal phase to be calculated for all standard formats. Essentially the use of PTP clocks on an IP network replace the need for both time-of-day timecode distribution (e.g. LTC) and other reference signals (e.g. BlackBurst).

'REAL-TIME' PROCESSING

The early days of television broadcasting certainly relied on 'real-time' processing. The scanning electron beam in the tube of the 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.

Whilst signal flows through digital television studios, master control and transmission systems are considered to be real-time, in reality there are many places where signals are artificially delayed ensuring the 'real-time' behaviour 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.

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 the overly simplified figure 1 below shows).

Figure 1 – Audio spill through open talkback can be problematic if delayed

The move to IP doesn't directly imply an increase in processing latency; where native uncompressed IP interfaces and non-blocking 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, whilst live production will continue to be considered a 'real-time' 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 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 centre 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.

The main driver for remote production is to reduce costs, the priority being a reduction in people on-site 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 PTZ cameras or mounts with recall of preset positions may further reduce onsite effort.

With all remote production architecures there is an obvious risk factor which may force a broadcaster to continue with a traditional outside broadcast operation; When prodcued locally (on-site) 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, jeopardising revenues and is likely to have a negative impact to a broadcaster's reputation.

ARCHITECTURE 1: 'PRODUCTION WITH REMOTE SOURCES'

The simplest form of remote production is arguably not remote production at all. Major sporting events at fixed venues have already justified the investment in dedicated fibre links specifically for use by broadcasters to return multiple feeds to production centres 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 centre to be used (see figure 2). Importantly this means staff get to go home at night and potentially can work on coverage of multiple different events on the same day.
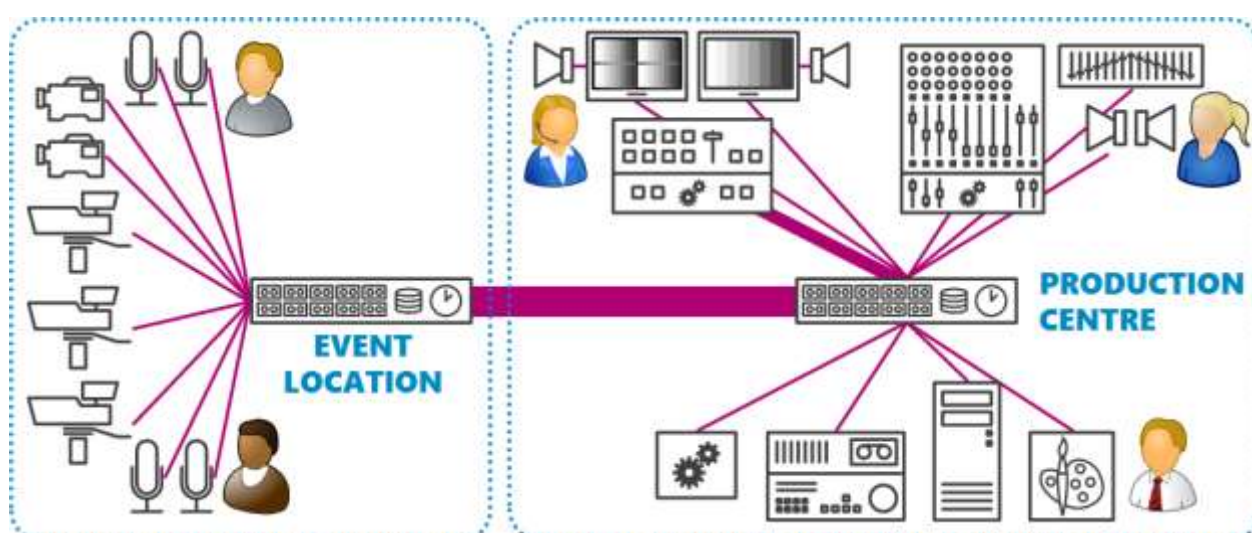

Figure 2 – Production with Remote Sources

This architecutre 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 centre.

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 feed back 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 centre 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 2 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:

1. Source Multiviewer
2. Mixer Program Output

Even with only light compression both these feeds would suffer some amount of delay. Figure 3 below shows a best-case example where sources are delayed by 4 frames (this could be more depending on how the source multiviewer feed is generated or depending on the compression and carriage mechanism used).
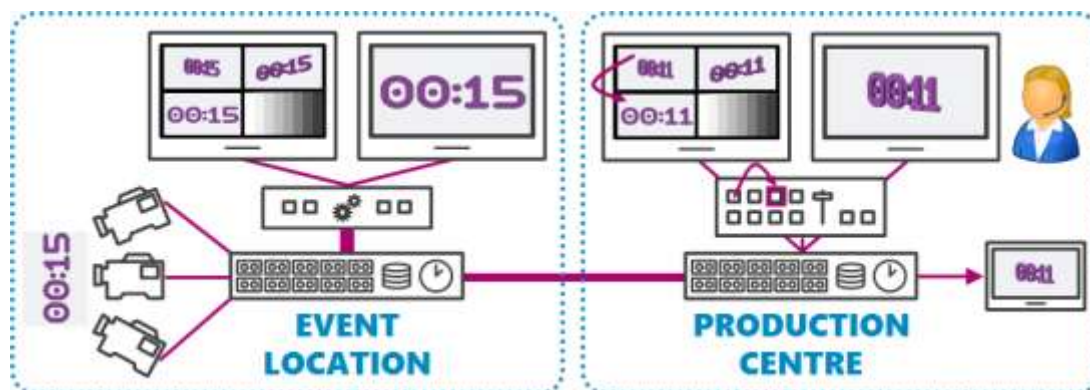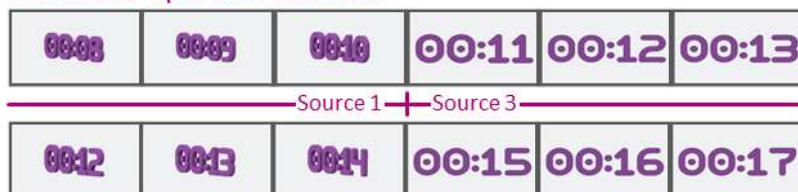

Figure 3 – Remote Controlled Vision Mixing

The example above highlights that whilst the equipment at the event location is operating in 'real-time' the operator's view will be showing frames in the past (e.g. Operator sees frame 00:11 whilst 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 4 frames later.

Figure 4 – The impact of monitoring latency on remotely controlled vision mixing

As figure 4 shows, 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 'real-time' way with all efforts placed on minimising 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 to them to where an operator is sat.

At the control location, which would typically be at the production centre but equally could be anywhere there is suitable IP connectivity, operational staff can view these proxies, which thanks to their timestamps, can be presented in a synchronous way. This method gives more flexibility for monitoring compared to a pre-compiled multiviewer feed.

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 which is unaffected by any jitter on the network.
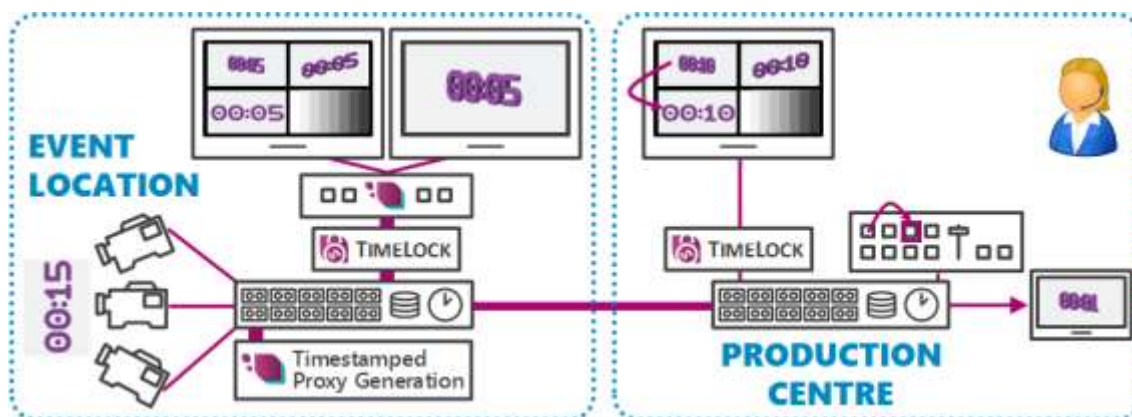

Figure 5 – Remotely controlled vision mixing with a fixed processing offset of 5 frames

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 second if responsive direction is required. In figure 5 above, the operator's monitoring is shown running with a 5 frame offset (i.e. the pictures the operator sees were captured 5 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. In figure 5 the full-res mixer is shown running at a delay of 10 frames – this the product of the 5 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 to be done).

This architecture can be visualised as a simple delay being applied to signals feeding into a mixer/switcher, in a software-based solution this can all be managed automatically.

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 (e.g. returned over an IP link) as they would have to wait an unacceptable time to see the effect of their actions.
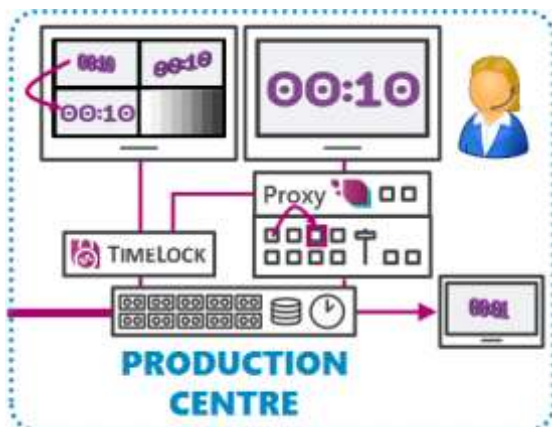

Figure 6 – Simulated mix process using proxy sources

To provide a better experience for the operator the timestamped proxy feeds can be used to perform a 'real-time' mix locally at the production centre, providing a simulation of what the onsite mixer will be doing slightly later (see figure 6).

With the current generation of broadcast hardware this proxy mixer setup may not be achievable but with software-based processing it is achievable with minimal additional hardware. A similar process using the proxy sources could also be used to generate a preview output to provide the full program/preset behaviour expected by vision operators (removing the need to return a preview feed from the onsite mixer).

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. 3rd party graphics, 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 centre 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 same offset).
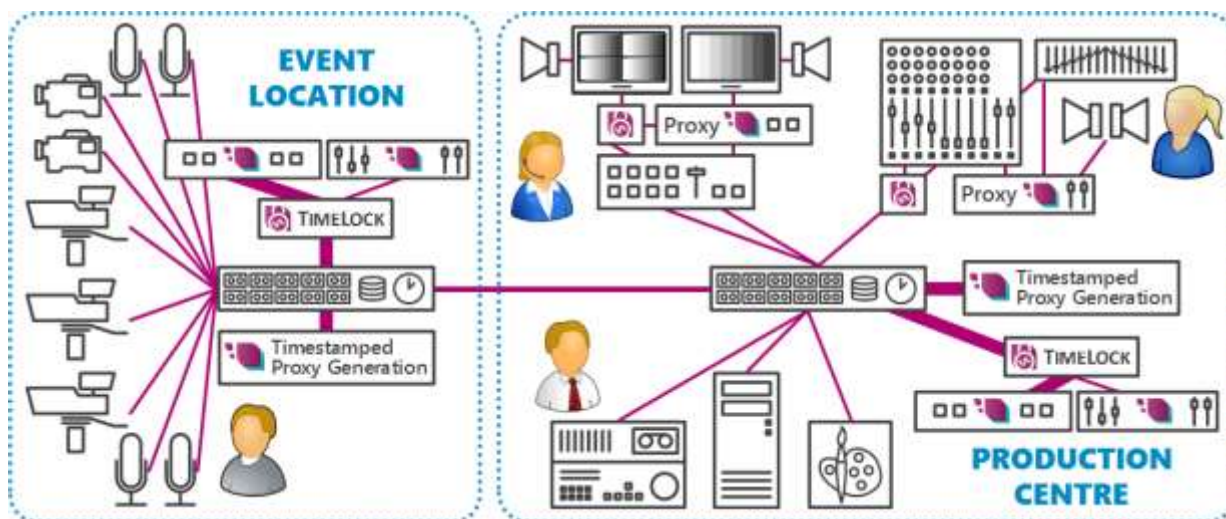

Figure 7 – Distributed remote production (multi-stage remotely controlled mixing)

A second set of full-resolution mixing would then be performed to mix between the local sources and the feed from the event location (see 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 architure described in figure 7 with specific signal architecture as detailed below in figure 8.

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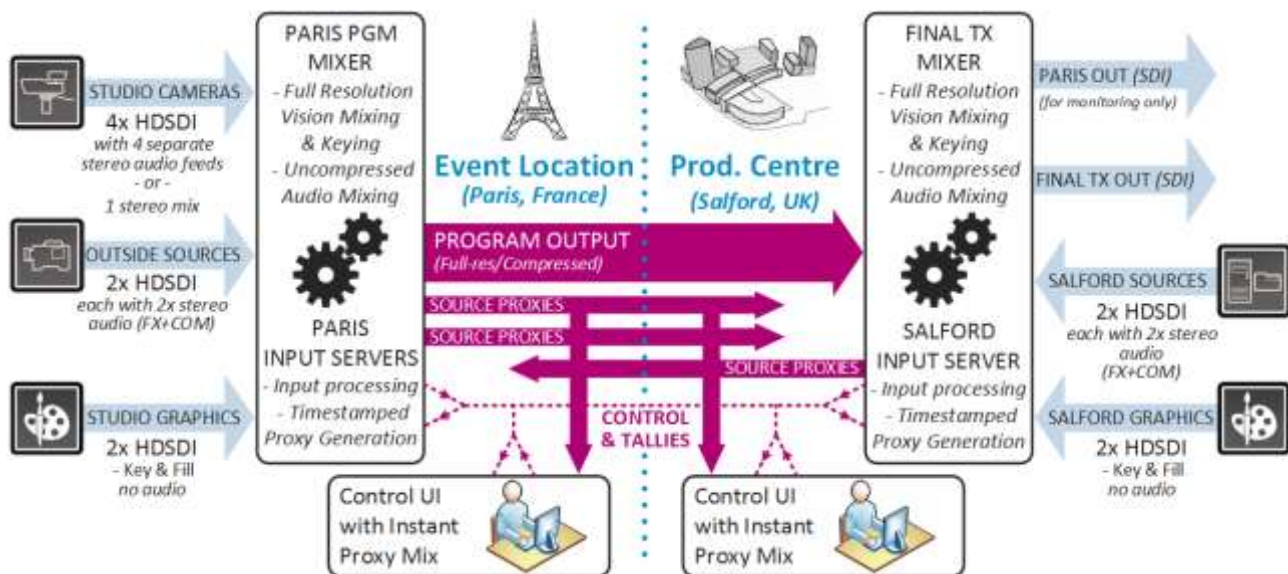

Figure 8 – Distributed production at Euro 2016 (multi-stage remotely controlled mixing)

At the production centre, a second software-based vision mixing process was run which switched between the feed from Paris and locally originated sources. The sources from the production centre 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 source also enabled a simulated mix to be generated in 'real-time' (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 50Mb/s.

CONCLUSION

Remote production should not be underestimated as simply being 'remote control', it requires methods for handling video and audio mixing of sources originate at the production centre as well as the event location. By compensating for latencies, distributed processing using multi-stage mixing can deliver viable architectures providing significant cost reductions. This will provide opportunities for broadcasters to consider televising events which would be uneconomical with traditional outside broadcast methods.