

Live Captions for Streaming

Hewson Maxwell

Head of Technology, Access Services - Red Bee Media

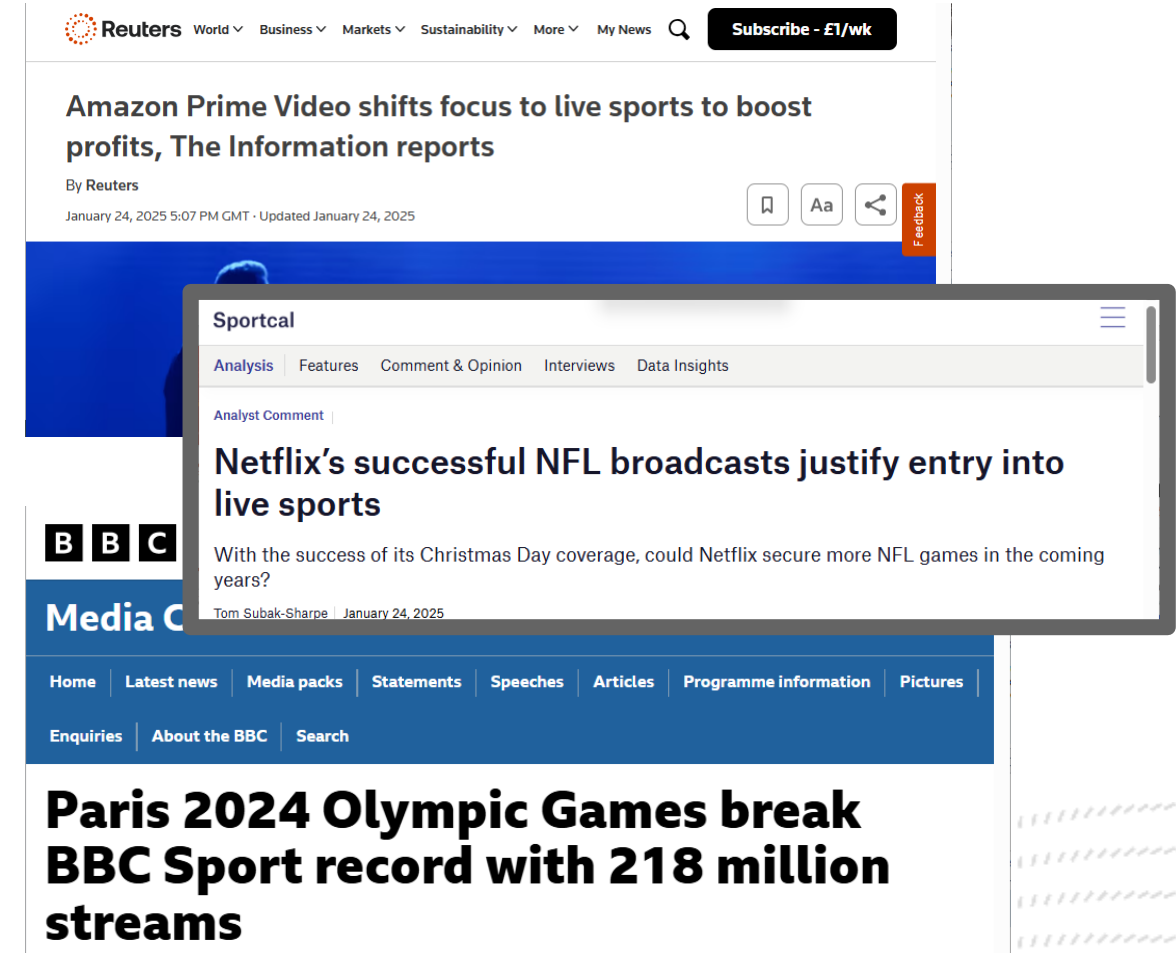
BORN IN BROADCAST, LIVE FOR MEDIA

COMING TO YOU LIVE... ON STREAMING



The Growth of Streaming-First Live Broadcast

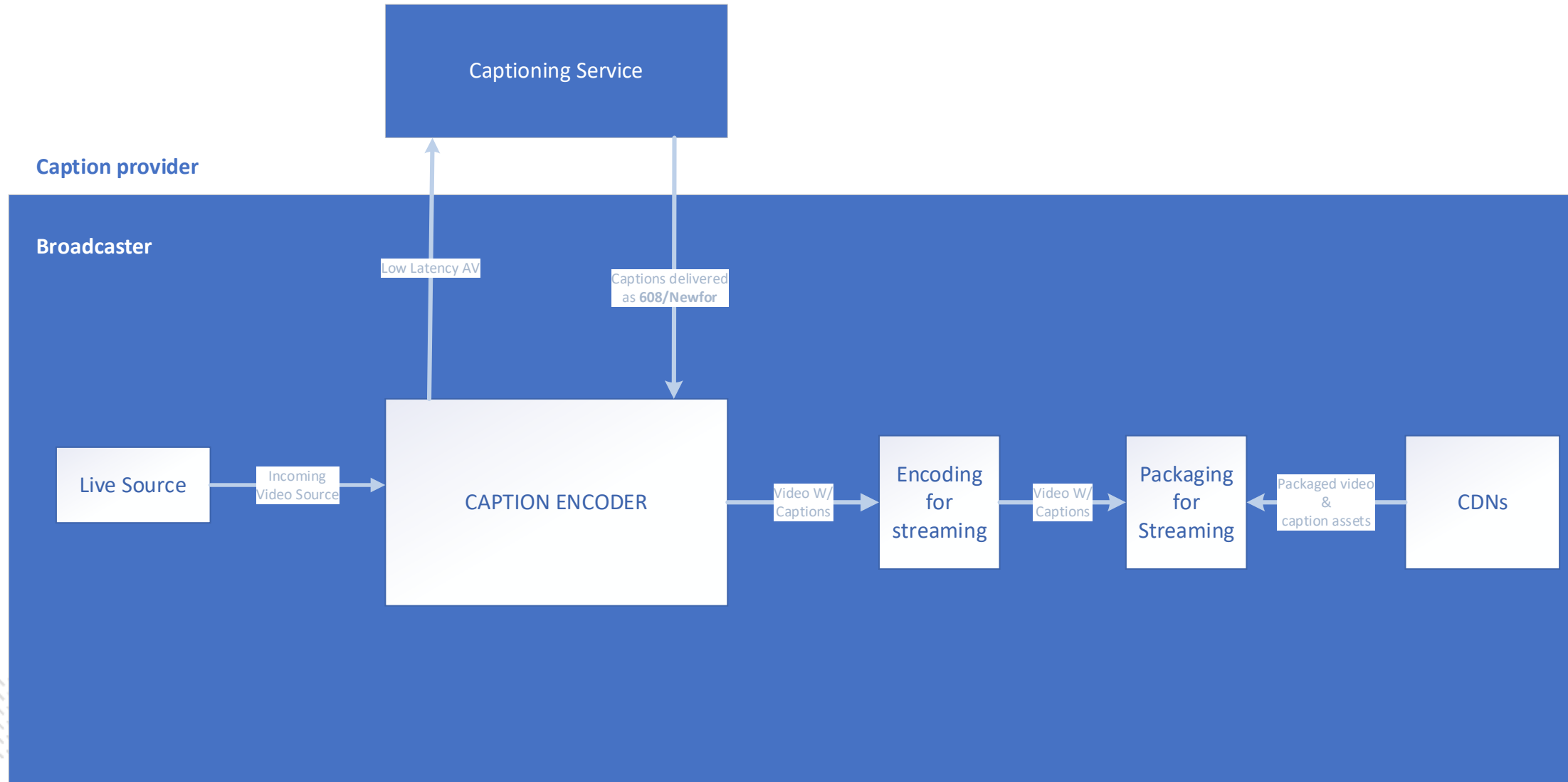
- An ever-increasing proportion of live broadcast is streaming-first
- This is true not only for streamers, but also for live broadcasters
- Cloud-based pop-up workflows allow the infrastructure to be brought up where required, allowing cost scaling and high levels of simultaneity
- For live captioning, various options are emerging for making live streams accessible, with different pros and cons



TRIED AND TRUSTED



The Legacy Approach



WHY CHANGE A WINNING FORMULA?



Legacy Pros & Cons

PROS

- Easy to use in workflows that support both linear & streaming
- Plugs and plays with high proportion of existing workflows for ease of deployment
- 608/708 / OP-47 widely carried by hardware and software solutions
- Known quantity for caption suppliers and live operations

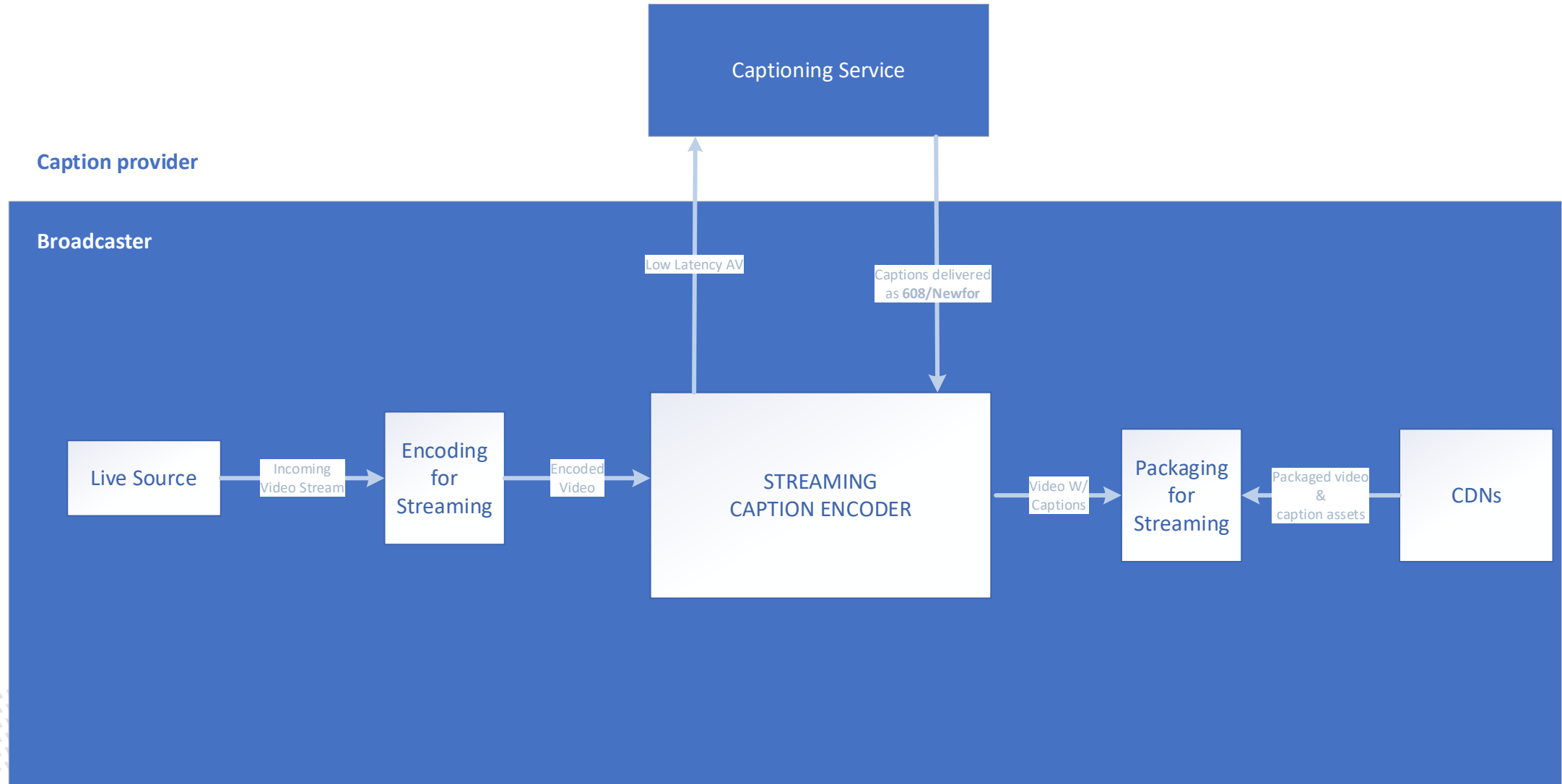
CONS

- Usually involves caption latency vs spoken word unless adding signal delay
- Limited character set support
- Limited support for multiple parallel languages
- Limited presentational options and control

ALTERNATIVE APPROACHES



Native Encoding for Streaming



VIRTUALISING ENCODING



Streaming Native Caption Encoders

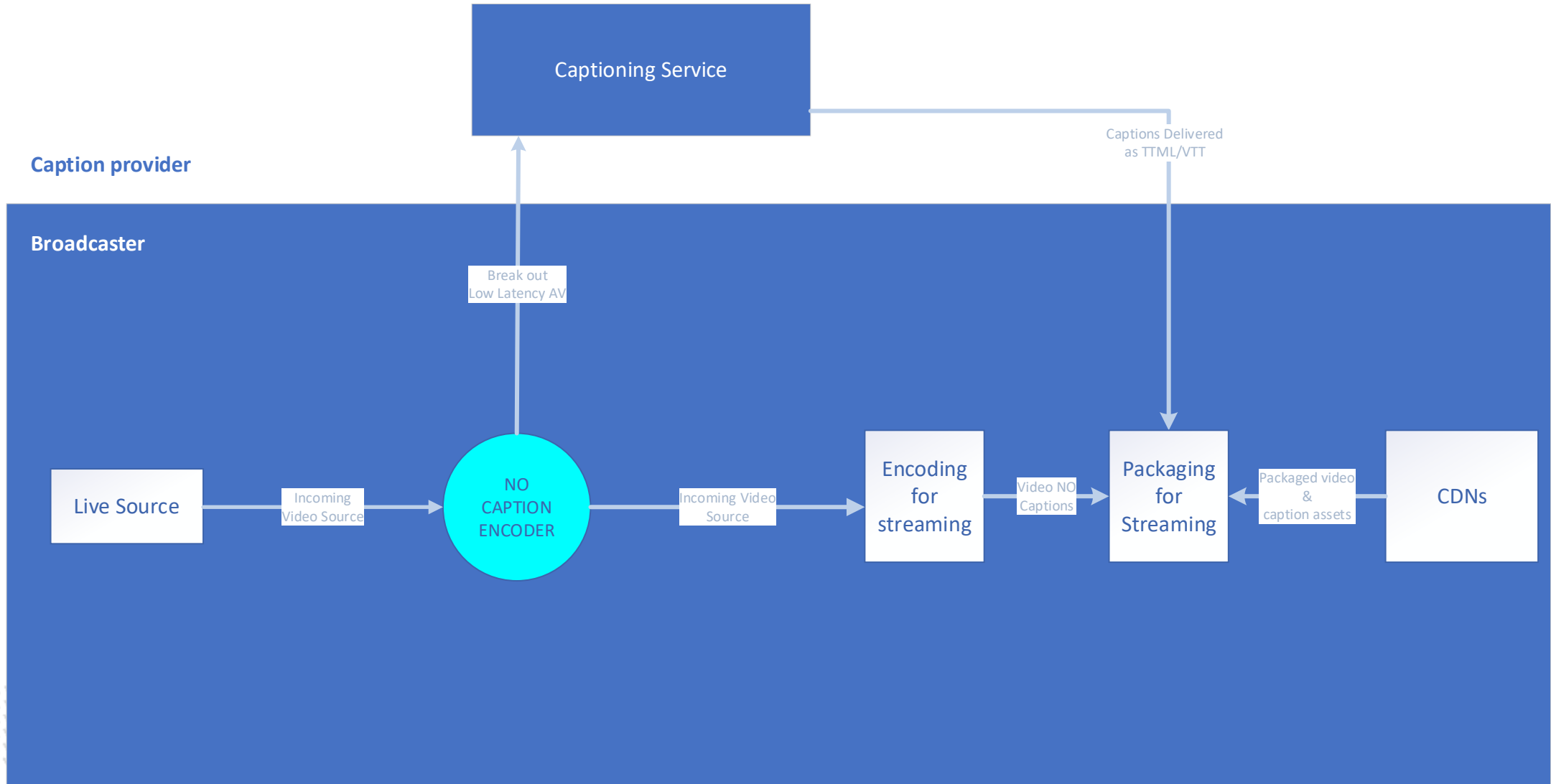
- Encoded video (via SRT, Zixi, etc) for signal acquisition and delivery
- Cloud deployed, so highly scalable.
- Wider subtitle encoding support for additional and simultaneous languages (e.g. multiple languages of auto-translated DVB-TTML via SyncWords)
- Option of adding delay to resynchronise – but it adds delay to overall delivery.



BEYOND THE CAPTION ENCODER



No Encoder Workflows



BEYOND THE CAPTION ENCODER



General approach – Pros & cons

PROS

- Allows for low or zero latency caption presentation without introducing additional signal delay
- Gives full Unicode language support
- Allows for many simultaneous languages or language-options to be delivered
- Gives a wider array of presentation options and control

CONS

- Lack of widely adopted standardised approaches mean cost and complexity for streamer
 - And for the caption provider/solution
- Parallel linear delivery requires different caption workflow
- Need to coordinate with different areas of responsibility within broadcast organisation



LIVE CAPTION DELIVERY



How are captions to be delivered?

CUSTOM INTERFACES

- Some organisations have defined HTTP interfaces for delivery of live captions
- Each caption segment delivered as a new document via HTTP PUT/POST
 - Segments can be event based, or cover all events within an epoch
 - Timing and ordering need to be managed
- Payload usually TTML (IMSC/EBU-TT) Based
- Works well, but not easily reusable by other organisations

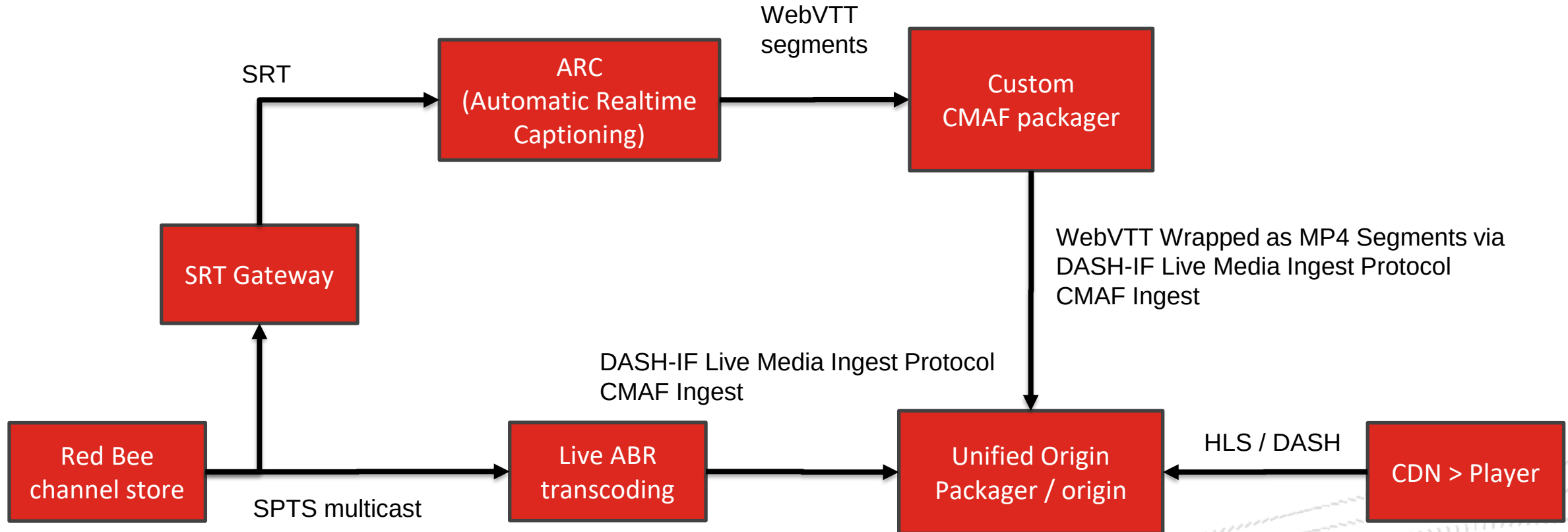
A STANDARDS BASED APPROACH?

- At MWS, Red Bee have been demoing a PoC making use of the DASH-IF Media Ingest protocol
- WebVTT segments in MP4 wrapper delivered via HTTP PUT/POST
- A simpler and cleaner approach would be delivering TTML or WebVTT documents directly
- A simple standard supported by packagers & encoders would be simpler for wider adoption by a variety of broadcast & captioning organisations

DASH-IF Live Media Ingest Protocol



Standards-based approach PoC



LATENCY CONSIDERATIONS



- Latency vs spoken word is one of the biggest drivers of caption usability for users
 - From receipt of source speech audio, high quality text takes at least 3-4 seconds to produce
- Translated subtitles take even longer to produce reliably. (The longer the better!)
- Where the latency between point of caption AV acquisition and caption delivery point is > 3-4 seconds, you can give captioning & translation subtitling more time to get things right
- Where and How to pick up the source stream influences the size of the window above, and the ease of integration with captioning solutions
- The emergence of low latency streaming is great! But it reduces the window for zero latency captions.



CONCLUSION



- Better and more usable captioning for audiences is good for everyone!
- Latency is one of the biggest drivers of live caption usability for audiences
- Organisations with smaller budgets need simple standardised approaches
- What can we do as a group to drive adoption & excellence?
 - Share approaches
 - Drive use of standards
 - Standards refinement

