SACJ 38(2) July 2026
Research Article

# The Edge as Origin: Five Architectural Transitions Enabling Real-Time Media on Consumer ARM Platforms

Madhav Narayan Bhat

Independent Researcher, USA

**ABSTRACT**
Consumer premise devices have transformed in under a decade from one-way broadcast receivers into IP-native, cloud-integrated real-time media hubs capable of sub-second two-way video, remote home streaming, and cross-vendor smart-home interoperability. This article analyzes that transformation through five architectural case studies drawn from the consumer electronics industry: WebRTC enablement on embedded Linux set-top boxes; server-side HTTP Live Streaming (HLS) on resource-constrained home hardware; hardware-accelerated low-latency pipelines on ARM systems-on-chip (SoCs); smart-home media interoperability bridging televisions, doorbells, and camera ecosystems; and the evolution of adaptive bitrate streaming toward chunked and low-latency variants. Each case study is selected because it represents a discrete architectural transition—not an incremental improvement—in the role of the consumer edge device. The argument advanced is that the embedded edge has become the decisive locus of innovation in consumer real-time media and that this shift was made possible by a specific sequence of open-standard adoptions (WebRTC, HLS, CMAF, and Matter) that displaced proprietary stacks and moved origination capability from cloud infrastructure to the device in the home. This study examines the implications for accessibility, privacy, energy efficiency, and the next codec generation (AV1, VVC).
**Keywords:** WebRTC; HTTP Live Streaming; embedded systems; low-latency streaming; smart-home interoperability; edge media processing; ARM SoC; V4L2; adaptive bitrate streaming.

**Article history**
Received: 05 June 2026
Accepted: 05 July 2026
Online: 31 July 2026

## 1. INTRODUCTION

For most of television's history, the consumer premise device was a one-way receiver. A cable or satellite set-top box (STB) demodulated a downlink, decoded MPEG-2 or H.264, and rendered it to an HDMI output. By contrast, the consumer edge device of 2026 is expected to originate real-time video, serve HLS sessions to distant mobile clients, bridge to smart-home cameras over WebRTC, and interoperate with heterogeneous ecosystems through Matter. The scale of this transition becomes apparent through market data: on-demand streaming alone accounts for 54% of total downstream internet volume, according to Sandvine's 2024 Global Internet Phenomena Report, while Ericsson's Mobility Report placed video at roughly 75% of all mobile data traffic at the end of 2025 [1], [2]. Pay-TV has contracted in parallel—the largest U.S. operators lost approximately 5.0 million net video subscribers in 2023, and S&P Global Market Intelligence placed U.S. pay-TV household penetration at 34.4% at the end of 2024, down from above 80% in 2011 [3]. The device that remained commercially viable through this contraction did so by acquiring capabilities its original design did not anticipate.

Three macro-trends underlie that capability acquisition. The IETF and W3C jointly promoted WebRTC to RFC/Recommendation status on 26 January 2021 [4], and the W3C issued an updated

Bhat, M.N. (2026). The Edge as Origin: Five Architectural Transitions Enabling Real-Time Media on Consumer ARM Platforms. *South African Computer Journal* 38(2), xx–xx.

ISSN 1015-7999 (print) ISSN 2313-7835 (online)

WebRTC Recommendation in March 2025 [5]. HLS was formalized in RFC 8216 in 2017 [6], with low-latency extensions now tracked in draft-pantos-hls-rfc8216bis [7]. MPEG-DASH reached its fifth edition in ISO/IEC 23009-1:2022 [8], and the Common Media Application Format (CMAF) reached its third edition in 2024 [9]. Standards maturation is necessary but not sufficient: the decisive engineering question is whether standards-compliant, sub-second real-time media can run inside the power and cost envelope of a consumer edge device. Consumer STBs and Linux-class smart-home gateways are commonly built around low-power ARM Cortex-A SoCs. Older mid-range STBs such as Amlogic S905/S905X used quad Cortex-A53 designs [10], while newer S905D3-class parts use quad Cortex-A55 cores. These SoCs typically include fixed-function media engines that offload video codec work; for example, S905D3 specifies 4K60 decode and H.265/H.264 encode up to 1080p60. Such platforms are usually fanless or passively cooled in few-watt operating envelopes, with embedded-scale cache hierarchies and DRAM configurations ranging from hundreds of megabytes to a few gigabytes.

The five case studies that follow are selected precisely because each one resolves a distinct version of that question. Case Study 1 asks whether a broadcast-class STB can become a WebRTC endpoint. Case Study 2 asks whether the same device can serve HLS sessions as an origin server. Case Study 3 examines the codec infrastructure transition (OMX to V4L2) that made both feasible at scale. Case Study 4 traces how smart-home video interoperability moved from per-vendor WebRTC hacks to standardized Matter 1.5 clusters. Case Study 5 analyzes the algorithmic and packaging evolution that collapsed broadcast-style latency from 20-30 seconds toward 2-3 seconds. Taken together, the cases argue that the embedded edge has crossed a qualitative threshold: it is no longer a terminal for media delivery but an origin point for it.

The discussion is grounded in IETF and W3C standards, ISO/IEC codec and packaging specifications, peer-reviewed systems research, and analyst reporting. Table 1 summarizes the latency and scalability characteristics of the protocols examined.

| System | Latency (sec) | Scalability |
|---|---|---|
| Traditional OTT / Broadcast-style streaming / Conventional live delivery | 20-30 | High (CDN-scaled) |
| Stream-to-phone / Remote home streaming / Low-latency live events | 2-3 | High (CDN-scaled) |
| Video doorbell preview / Caller-to-TV / Two-way A/V | <0.5 | Low (peer-scaled) |

Table 1. Latency and scalability profile of embedded media delivery protocols examined in this article. Latency figures drawn from published benchmarks; see individual case studies for citations.

## 2. CASE STUDY 1 -- WEBRTC ON EMBEDDED LINUX SET-TOP BOXES

Compliant WebRTC endpoints must terminate ICE, DTLS 1.2+, SRTP, congestion-controlled RTP (RFC 8834), and SCTP-over-DTLS for data channels (RFC 8831), with signaling expressed through

Bhat, M.N. (2026). The Edge as Origin: Five Architectural Transitions Enabling Real-Time Media on Consumer ARM Platforms. South African Computer Journal 38(2), 77–86.

ISSN 1015-7999 (print) ISSN 2313-7835 (online)

JSEP offer/answer (RFC 9429) [4], [11], [12], [13]. On a browser, this protocol surface is Chromium's problem. On an embedded Linux STB running a quad Cortex-A53 with 512 MB of DRAM and a 5W passive cooling envelope, it is an architectural problem—one that Chromium's reference libwebrtc, with its substantial C++ footprint and dependency weight, does not solve for legacy SoC build systems.

Two architectural patterns have emerged. GStreamer's webrtcbin, introduced in GStreamer 1.14 (2018), implements much of the W3C RTCPeerConnection model and is designed to leverage GStreamer's existing hardware-accelerated encoding/decoding, zero-copy, and embedded-platform support [14]. Critically, a GStreamer-based design can integrate hardware-backed v4l2h264enc/v4l2h264dec elements rather than the software encoders bundled in libwebrtc, enabling DMA-mapped buffer handoff between capture, encode, and RTP packetization [14], [15]. The second pattern is pure-language stacks such as Pion (Go) and AWS's Amazon Kinesis Video Streams WebRTC SDK in C, the latter explicitly targeting doorbell and baby-monitor-class devices with a stated sub-one-second latency envelope [16]. The trade-off is well characterized: libwebrtc maximizes browser parity; native stacks yield smaller binaries and more flexible codec integration at the cost of feature drift against Chromium behavior.

Empirical results on embedded ARM targets are now solid. Nakagawa et al. measured WebRTC round-trip on Raspberry Pi 4-class hardware and found the protocol stack added only 1-3 ms above ICMP ping latency across enforced delays of 0, 10, and 100 ms confirming that the protocol overhead is not the bottleneck on modern ARM [17]. Practitioner testing of GStreamer-based edge pipelines has reported 142-159 ms glass-to-glass latency on NVIDIA Jetson-class SoCs, competitive with cloud SFUs [14]. The productionization burden is not latency but state management: bridging the STB's stateful broadcast A/V pipeline with WebRTC's bidirectional RTP session, negotiating TURN on the approximately 22% of sessions requiring relay [18], and sustaining thermal budgets during multi-hour sessions.

This case study demonstrates the first of the article's five transitions: the consumer edge device moves from a media terminal to a WebRTC endpoint, acquiring bidirectional real-time video capability without cloud relay for the majority of sessions.

## 3. CASE STUDY 2 SERVER-SIDE HLS ON EMBEDDED HARDWARE FOR HOME STREAMING

HLS, specified in RFC 8216 [6], was designed for CDN-fronted broadcast delivery, where load is distributed across thousands of edge points of presence. Repurposing it so that a set-top box inside a consumer home can serve authenticated multivariant playlists and fragmented MP4 segments to a mobile Safari client on the far side of the public internet inverts the scaling model entirely. A CDN-fronted HLS service distributes its load across tens of thousands of edge PoPs; the in-home origin has exactly one CPU and one upstream link. That constraint is the central engineering problem of this case study.

Two observations motivate the architecture nevertheless. Safari on iOS and macOS carries the most mature native HLS client in the industry, with hardware-accelerated H.264/HEVC decode and strict conformance to Apple's HLS Authoring Specification [19]—which eliminates the need for custom mobile clients. Because HLS uses ordinary HTTP GETs against short-lived segment URLs, it traverses NATs, captive portals, and carrier middleboxes with significantly less friction than WebRTC's ICE-negotiated UDP paths. A tier-one satellite or cable operator deploying a "stream-to-



ISSN 1015-7999 (print) ISSN 2313-7835 (online)

phone" feature can therefore support hundreds of thousands of concurrent remote viewers across heterogeneous mobile networks without the signaling or TURN overhead of WebRTC.

Three optimizations address the single-origin constraint. Low-latency HLS extensions: Partial Segments, Blocking Playlist Reload, preload hints, and rendition reports introduced by Apple at WWDC 2019 [20] and generalized at WWDC 2020 [21] reduce live latency from the traditional 20-30 seconds characteristic of 10-second-segment HLS to a target below 2 seconds [7], [22]. CMAF (ISO/IEC 23000-19) allows a single fragmented-MP4 encoding to feed both HLS and DASH clients, halving storage requirements on the embedded origin [9]. Chunked transfer encoding of CMAF fragments, as described in Law's 2018 The Akamai whitepaper releases frames to the wire as they are encoded, permitting sub-three-second end-to-end latency on Safari with no client-side protocol changes [22].

The second transition this case study represents is distinct from the first: where WebRTC made the edge device a real-time communications endpoint, server-side HLS makes it a broadcast origin server—serving tens of thousands of remote sessions from hardware designed for single-stream playout.

## 4. CASE STUDY 3 -- LOW-LATENCY MEDIA PIPELINES ON CONSTRAINED ARM HARDWARE

The OMX deprecation is the architectural inflection point that made the prior two case studies tractable at scale. Khronos's OpenMAX IL last saw a formal release in version 1.1.2 in September 2008; version 1.2 remained provisional since 2011 and was never ratified [23]. Raspberry Pi's deprecation of its OMX stack, followed by Firefox's addition of V4L2-M2M hardware decode in 2024 to match Chromium's ChromeOS implementation [24], completed a migration that had been underway since mainline kernel support for V4L2 stateful and stateless codec interfaces matured [15], [25]. Chromium/ChromeOS video acceleration documentation and design discussions now describe V4L2 stateful and stateless hardware-decode paths for ARM-class Linux platforms rather than the retired OMX plugin [26]. Fig. 1 depicts various features of the shift from classical to zero-copy WebRTC media.

The practical consequence for consumer devices is zero-copy pipeline architecture. A GStreamer pipeline on a mid-range STB SoC—a quad Cortex-A53/A55-class platform with a dedicated hardware encoder rated at up to 1080p60 H.264/H.265 on representative Amlogic parts—can chain v4l2src to v4l2h264enc to rtph264pay to webrtcbin with DMA buffer handoff between capture, encode, and RTP packetization stages, eliminating the memory copies that had historically made software-based pipelines thermally unacceptable in few-watt passive envelopes [10], [14], [15]. Because the V4L2 Request API permits per-frame control submission for stateless decoders, userspace can implement reference-frame management—essential for WebRTC's selective retransmission and simulcast layers—without kernel modifications.

Quantitative overhead matters because the edge budget is unforgiving. Peer-reviewed work on Raspberry Pi class ARMv7 devices shows that a hybrid Electron/React-Native WebRTC application consumes approximately 4% less CPU and 7% less RAM than full Chromium during a live call [27]. A tier-one consumer electronics OEM producing a novel embedded media feature at scale wins or loses on codec integration discipline: correct handling of hardware-encoder bitrate-controller semantics, careful pairing of V4L2 buffer counts with network pacing, and instrumentation of RTCP feedback loops to drive encoder bitrate targets dynamically.

Bhat, M.N. (2026). The Edge as Origin: Five Architectural Transitions Enabling Real-Time Media on Consumer ARM Platforms. South African Computer Journal 38(2), 77–86.

ISSN 1015-7999 (print) ISSN 2313-7835 (online)

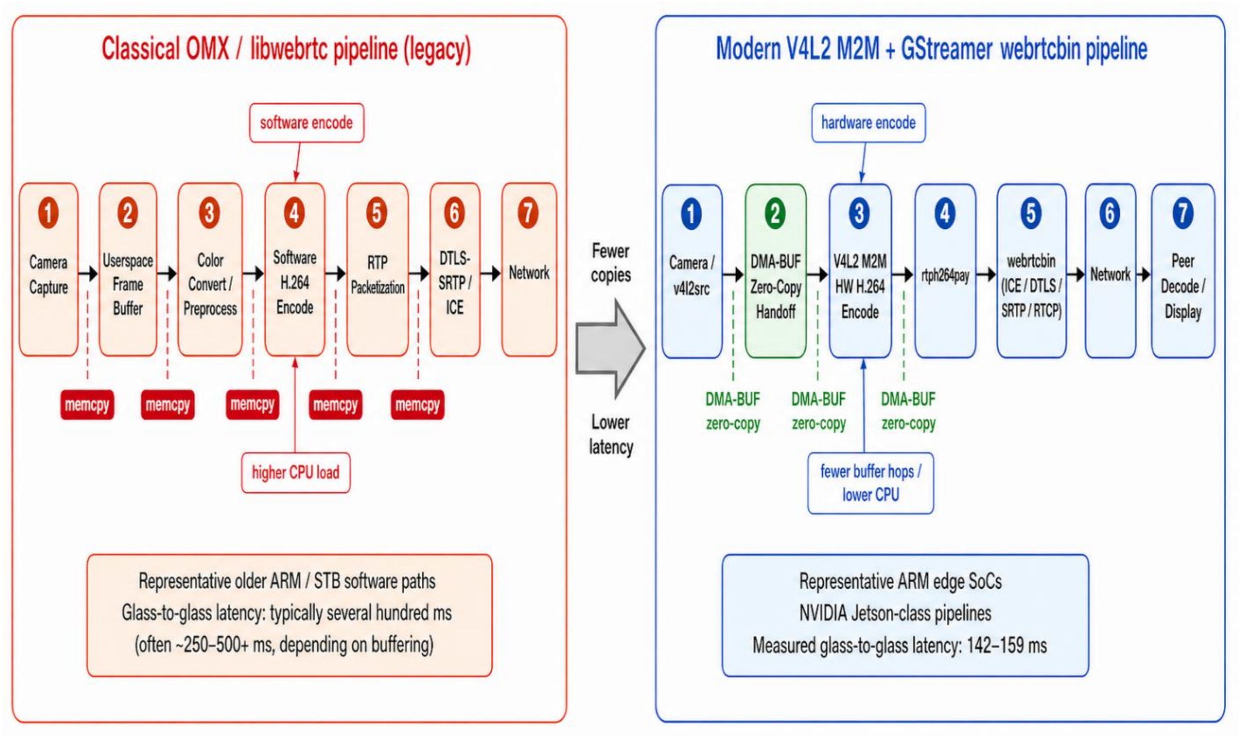


Figure 1. Classical versus zero-copy WebRTC media pipelines on ARM SoCs.

The V4L2 transition represents the third of the five architectural shifts: an infrastructure-level change that lowered the implementation cost of Cases 1 and 2 by an order of magnitude, enabling them to move from prototype to mass production at consumer scale.

## 5. CASE STUDY 4 -- SMART-HOME MEDIA INTEROPERABILITY

The Connectivity Standards Alliance published Matter 1.0 in October 2022 covering lighting, HVAC, locks, and media devices—but pointedly excluded cameras [28]. The omission was not an oversight; it reflected unresolved technical and commercial disagreements about live-stream transport, privacy controls, and codec standardization. The three-year gap it created was commercially significant: consumer platforms that wanted to surface a video-doorbell preview on the living-room television had to implement a separate pipeline per partner camera manufacturer, each using WebRTC as an ad-hoc transport and vendor-specific REST APIs for signaling. Several major operators built and shipped these integrations, received industry recognition for them, and then faced the prospect of replacing them with a standardized equivalent.

Matter 1.5, published on 20 November 2025, introduced Camera, Video Doorbell, Snapshot Camera, and Intercom device types and defined WebRTC (with STUN/TURN) as the canonical transport for live A/V streaming, signaled through new "WebRTC Transport Provider/Requestor," "Camera AV Settings," and "Zone Management" clusters [29]. The market response was immediate: Samsung SmartThings announced the first Matter 1.5 camera ecosystem in 2025, integrating third-party doorbells and cameras with live streaming, two-way audio, motion detection, and pan/tilt/zoom via Matter clusters [30].

The market scale justifies the consolidation. Parks Associates reports that 30% of U.S. internet households owned either a smart camera or smart video doorbell in 2024, averaging 2.21 video devices per owning household [31]; video-doorbell household penetration grew from 4% in 2017 to



ISSN 1015-7999 (print) ISSN 2313-7835 (online)

roughly 20% by 2023 [32]. Grand View Research values the global doorbell-camera market at USD 2.02 billion in 2024, projected to USD 6.84 billion by 2033 [33]. With WebRTC now the Matter streaming transport, the set-top box class of device is positioned to become the home's universal video-hub endpoint, terminating WebRTC sessions from any Matter-compliant camera regardless of vendor.

The fourth transition is standardization: what had been a collection of bilateral integrations between specific operators and specific camera manufacturers becomes a single cluster-based interface, lowering integration cost and enabling any Matter-compliant edge device to serve as the hub.

## 6. CASE STUDY 5 -- ADAPTIVE STREAMING EVOLUTION AND LOW-LATENCY HLS

Three algorithmic generations now coexist in deployed ABR clients, and examining their sequence reveals the underlying logic of the field's progress. Buffer- and throughput-based heuristics such as BOLA apply Lyapunov optimization and are provably within O(1/V) of offline optimal; BOLA is the default in dash.js and is deployed at Akamai, the BBC, and major broadcasters [34]. Its limitation is that it treats throughput as a stationary signal, which it is not on mobile networks. Model-predictive control approaches—notably the MPC/RobustMPC family from Yin et al.—address this by combining throughput estimation with explicit predictive control over a finite horizon [35], achieving better quality consistency under non-stationary conditions at the cost of model accuracy assumptions. Reinforcement-learning approaches, of which Pensieve is canonical, bypass the modeling assumption entirely by learning policies directly from observed network traces; Pensieve outperforms prior work by 12-25% on composite QoE metrics [36]. Oboe further auto-tunes BOLA/MPC parameters per network state [37].

Packaging evolution has advanced in parallel and is arguably the more consequential development for embedded origins. The industry has collapsed toward CMAF chunked encoding as the common substrate for both HLS and DASH low-latency delivery. Akamai's reference architecture demonstrated that a 4-second segment encoded at 30 fps with one frame per chunk releases content approximately 3.97 seconds earlier than waiting for full-segment delivery, with an end-to-end latency of approximately 2.8 seconds at 4.5 Mbps 1080p AVC [22]. Sub-second latency is achievable under aggressive chunking. This closes much of the gap to WebRTC for broadcast-style use cases while preserving HTTP's cacheability.

Industry telemetry confirms the adoption trajectory. Bitmovin's 2024/2025 Video Developer Report, surveying 167 developers across 34 countries, reports the low-latency technology mix at LL-HLS 34.8%, LL-DASH 23.9%, and WebRTC 22.8% [38]. Conviva's State of Streaming reports provide large-scale operational telemetry—embedded in roughly 3.3 billion streaming applications and measuring about 180 billion streams annually in the 2021 methodology, later reported at 4.0 billion applications and 200 billion streams annually in 2022—and document buffering, startup-time, and bitrate trends as directional quality-of-experience indicators [39].

The fifth transition completes the argument: adaptive streaming is no longer a cloud-only intelligence function. The ML-based ABR policy, the chunked CMAF encoder, and the LL-HLS origin all run—or can run—on the ARM SoC inside the consumer premise device, closing the latency gap to WebRTC for linear use cases without the session-management complexity WebRTC imposes.

Bhat, M.N. (2026). The Edge as Origin: Five Architectural Transitions Enabling Real-Time Media on Consumer ARM Platforms. South African Computer Journal 38(2), 77–86.

ISSN 1015-7999 (print) ISSN 2313-7835 (online)

## 7. Broader Industry and Societal Implications

The compound effect of these five architectural transitions extends well beyond entertainment. Telehealth provides the clearest illustration: government reporting shows Medicare telehealth services rose from roughly 5 million in April-December 2019 to more than 53 million in the same months of 2020, and behavioral-health and psychiatry video visits continued growing even as general telehealth moderated post-COVID [40], [41]. Real-time video at the edge—particularly on large-screen devices in the living rooms of elderly or mobility-limited users—converts what was once a smartphone-only modality into a household-scale access point. The WebRTC-enabled STB with a paired camera effectively becomes a universal communications terminal available to users who cannot or do not operate smartphones.

The regulatory and privacy surface has grown in step with capability. The U.S. FCC's Twenty-First Century Communications and Video Accessibility Act (CVAA) of 2010 requires interoperable video conferencing and IP-delivered video programming to support captioning and audio description pass-through [42]—obligations that directly reach embedded WebRTC and HLS origins. In the EU, the Court of Justice's Rynes ruling (C-212/13, 2014) established that home cameras capturing any public space fall outside the GDPR household exemption [43], a precedent that shapes how video-doorbell vendors must design retention, encryption, and access-control defaults. Engineers who deploy WebRTC or Matter streaming to production consumer devices inherit these obligations whether or not they were considered at design time.

The energy dimension is increasingly material. Peer-reviewed modeling indicates that edge-resident media processing can reduce energy consumption by approximately 19-28% compared with equivalent cloud processing, chiefly by avoiding transmission energy [44]. Preist, Schien, and Shabajee's CHI 2019 study estimated global YouTube viewing emissions at roughly 10 million tonnes of $CO_2$ in 2016, with mobile networks as the largest contributor [45]. As streaming and real-time video continue to absorb an increasing share of internet traffic, edge-resident media processing is likely to become not just a quality-of-experience choice but an energy-policy one.

## 8. LIMITATIONS

The five case studies examined in this article are selected from the consumer set-top box and smart-home sectors of the embedded media landscape. Automotive in-cabin media, industrial video monitoring, and broadcast production systems share some of the architectural patterns discussed but operate under sufficiently different regulatory, latency, and reliability constraints that the analysis should not be applied to them without domain-specific adaptation. Quantitative benchmarks—latency figures, throughput measurements, and rebuffering rates—are drawn from published literature, practitioner reports, and standards documents; they reflect hardware generations and network conditions at the time of measurement and should be validated against current SoC and network profiles before being used prescriptively. Market forecasts from analyst firms are cited as directional indicators rather than precise projections. The energy-efficiency estimates cited in Section 7 include a preprint source alongside a peer-reviewed Elsevier publication; the former should be treated as preliminary pending journal review.

## 9. CONCLUSIONS

The five case studies trace a single architectural argument: the consumer premise device has undergone a qualitative transition from media terminal to media origin, and that transition was enabled by a specific sequence of events—WebRTC standardization, HLS's low-latency extensions,

Bhat, M.N. (2026). The Edge as Origin: Five Architectural Transitions Enabling Real-Time Media on Consumer ARM Platforms. South African Computer Journal 38(2), 77–86.

ISSN 1015-7999 (print) ISSN 2313-7835 (online)

the OMX-to-V4L2 migration, Matter 1.5's camera clusters, and ML-based ABR -- each of which moved a capability that had previously required cloud infrastructure to the device at the edge. The sequence is not coincidental. Each transition depended on the one before it: zero-copy V4L2 pipelines made WebRTC thermally viable on 5W STBs; chunked CMAF made LL-HLS viable from a single-origin CPU; Matter 1.5's adoption of WebRTC as its camera transport was possible because WebRTC had already been demonstrated as a viable embedded-device protocol.

What comes next is already visible. AV1, finalized by the Alliance for Open Media in 2018, is entering consumer hardware decode pipelines and will pressure HEVC licensing regimes [46]. VVC (H.266), finalized in July 2020 by ITU-T/ISO/IEC, promises roughly 50% bitrate savings over HEVC and is moving into broadcast deployment pathways such as Brazil's TV 3.0 field tests scheduled from 2025 [47]. WebRTC-NV work at the W3C is unlocking insertable streams, scalable video coding, and encoded-transform APIs that move more control to application logic without revisiting libwebrtc internals. End-edge-cloud collaborative inference—surveyed in IEEE COMST 2023 and 2024—points toward hybrid pipelines where the STB handles capture, encoding, and low-latency transport while cloud partners handle super-resolution, indexing, or accessibility enhancement [48], [49].

The broader implication is that the embedded edge has become the site at which standards compliance, codec physics, thermal constraints, and user experience converge. A decade ago, "real-time media" meant a dedicated video conferencing terminal or a cloud-rendered stream delivered to a passive receiver. It now means a consumer device manufactured at the cost of a set-top box, running open-standard protocols across a 5W power budget, capable of both originating and terminating bidirectional video. The engineering decisions that made this possible—open kernel APIs, standards-based protocol stacks, and chunked packaging formats—are the ones that will determine how the next decade of real-time media innovation is built.

## REFERENCES

[1] Sandvine, Global Internet Phenomena Report 2024, Mar. 2024.

[2] Ericsson, Ericsson Mobility Report: Mobile Traffic Update, June 2026, 2026.

[3] Leichtman Research Group, "Major pay-TV providers lost about 5,035,000 subscribers in 2023," press release, Mar. 8, 2024; T. Lafayette, "Pay TV Penetration Fell to 34.4% in Q4: S&P Global," TV Tech, Mar. 5, 2025.

[4] H. Alvestrand, "Overview: Real-Time Protocols for Browser-Based Applications," IETF RFC 8825, Jan. 2021, doi:10.17487/RFC8825.

[5] W3C, WebRTC: Real-Time Communication in Browsers, W3C Recommendation, Mar. 13, 2025.

[6] R. Pantos and W. May, "HTTP Live Streaming," IETF RFC 8216, Aug. 2017, doi:10.17487/RFC8216.

[7] R. Pantos, Ed., "HTTP Live Streaming 2nd Edition," Internet-Draft draft-pantos-hls-rfc8216bis-22, IETF, May 1, 2026.

[8] ISO/IEC 23009-1:2022, Information Technology -- Dynamic Adaptive Streaming over HTTP (DASH) -- Part 1: Media Presentation Description and Segment Formats, 5th ed., Aug. 2022.

[9] ISO/IEC 23000-19:2024, Information Technology -- Multimedia Application Format (MPEG-A) -- Part 19: Common Media Application Format (CMAF) for Segmented Media, 3rd ed., Feb. 2024.

[10] Amlogic, S905D3 Quick Reference Manual, rev. 0.6, Jul. 2019; Amlogic, S905X Datasheet, rev. 0.3, Mar. 2017.

[11] C. Perkins, M. Westerlund, and J. Ott, "Media Transport and Use of RTP in WebRTC," IETF RFC 8834, Jan. 2021, doi:10.17487/RFC8834.

Bhat, M.N. (2026). The Edge as Origin: Five Architectural Transitions Enabling Real-Time Media on Consumer ARM Platforms. South African Computer Journal 38(2), 77–86.

ISSN 1015-7999 (print) ISSN 2313-7835 (online)

[12] R. Jesup, S. Loreto, and M. Tuxen, "WebRTC Data Channels," IETF RFC 8831, Jan. 2021, doi:10.17487/RFC8831.
[13] J. Uberti, C. Jennings, and E. Rescorla, Ed., "JavaScript Session Establishment Protocol (JSEP)," IETF RFC 9429, Apr. 2024, doi:10.17487/RFC9429.
[14] GStreamer Project, "GStreamer 1.14 Release Notes" and webrtcbin documentation, 2018; GStreamer Project, gst-plugins-good/sys/v4l2/gstv4l2h264enc.c, open-source reference implementation; RidgeRun, "GstWebRTC - Real Time Streaming with WebRTC," 2021.
[15] The Linux Kernel Community, "Video Memory-To-Memory Interface," Linux Kernel userspace-api documentation.
[16] Amazon Web Services, Amazon Kinesis Video Streams with WebRTC SDK in C Developer Guide, 2019-2024.
[17] H. Nakagawa et al., "WebRTC-based measurement tool for peer-to-peer applications and preliminary findings with real users," in Proc. ACM AINTEC, 2021, doi:10.1145/3497777.3498544.
[18] C. Hart, "The Big Churn - learning from real usage stats (Lasse Lumiaho and Varun Singh)," webrtcHacks, Apr. 8, 2016.
[19] Apple Inc., HLS Authoring Specification for Apple Devices, Apple Developer documentation, 2025.
[20] R. Pantos, "Introducing Low-Latency HLS," WWDC 2019 Session 502, Apple Inc., Jun. 2019.
[21] R. Pantos, "What's New in Low-Latency HLS," WWDC 2020 Session 10228, Apple Inc., Jun. 2020.
[22] W. Law, "Ultra-Low-Latency Streaming Using Chunked-Encoded and Chunked-Transferred CMAF," Akamai Technologies white paper, 2018.
[23] Khronos Group, OpenMAX Integration Layer Specification v1.1.2, Sep. 2008; v1.2 Provisional, Nov. 2011.
[24] Mozilla, "Bug 1833354 -- [Linux] Implement V4L2-M2M hardware-accelerated video decode," Mozilla Bugzilla, 2023-2024.
[25] The Linux Kernel Community, "Stateless Codec Control Reference," Linux V4L2 API documentation.
[26] The Chromium Project, "Linux HW Video Decode Overview" and V4L2 video-decode design/source documentation, 2022-2024.
[27] B. Diallo, A. Ouamri, and M. Keche, "A Hybrid Approach for WebRTC Video Streaming on Resource-Constrained Devices," Electronics, vol. 12, no. 18, art. 3775, 2023, doi:10.3390/electronics12183775.
[28] Connectivity Standards Alliance, "Matter 1.0 Arrives," press release, Oct. 4, 2022.
[29] Connectivity Standards Alliance, "Matter 1.5 Introduces Cameras, Closures, and Enhanced Energy Management Capabilities," press release, Nov. 20, 2025.
[30] Samsung Electronics, "Samsung SmartThings Becomes the Industry's First to Support Matter Cameras," newsroom release, 2025.
[31] Parks Associates, "30% of Internet Households Own Either a Smart Camera or a Smart Video Doorbell," press release, Apr. 30, 2024.
[32] Parks Associates, "Adoption of Video Doorbells Reaches to 20% of U.S. Internet Households in 2023," press release, 2023.
[33] Grand View Research, Doorbell Camera Market Size, Share, Industry Report 2033, 2024.
[34] K. Spiteri, R. Urgaonkar, and R. K. Sitaraman, "BOLA: Near-Optimal Bitrate Adaptation for Online Videos," IEEE/ACM Trans. Networking, vol. 28, no. 4, pp. 1698-1711, 2020, doi:10.1109/TNET.2020.2996964.



ISSN 1015-7999 (print) ISSN 2313-7835 (online)

[35] X. Yin, A. Jindal, V. Sekar, and B. Sinopoli, "A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP," in Proc. ACM SIGCOMM, 2015, doi:10.1145/2785956.2787486.

[36] H. Mao, R. Netravali, and M. Alizadeh, "Neural Adaptive Video Streaming with Pensieve," in Proc. ACM SIGCOMM, 2017, doi:10.1145/3098822.3098843.

[37] Z. Akhtar et al., "Oboe: Auto-tuning Video ABR Algorithms to Network Conditions," in Proc. ACM SIGCOMM, 2018, pp. 44-58, doi:10.1145/3230543.3230558.

[38] Bitmovin, The 8th Annual Bitmovin Video Developer Report 2024/2025, 2025. [Practitioner survey; cited as industry data.]

[39] Conviva, State of Streaming quarterly reports, Q3 2020-Q2 2022.

[40] U.S. Government Accountability Office, Medicare Telehealth: Actions Needed to Strengthen Oversight and Help Providers Educate Patients on Privacy and Security Risks, GAO-22-104454, Sept. 26, 2022.

[41] S. Mandal, B. Wiesenfeld, and O. Nov, "The 'New' New Normal: Changes in Telemedicine Utilization Since COVID-19," Am. J. Manag. Care, vol. 31, no. 3, pp. e74-e78, 2025, doi:10.37765/ajmc.2025.89700.

[42] U.S. Federal Communications Commission, Twenty-First Century Communications and Video Accessibility Act (CVAA), Public Law 111-260, Oct. 8, 2010.

[43] Court of Justice of the European Union, Frantisek Rynes v Urad pro ochranu osobnich udaju, Case C-212/13, Dec. 11, 2014.

[44] T. Huang, W. Huang, B. Zhang, W. Chen, and X. Pan, "Optimizing energy consumption in centralized and distributed cloud architectures with a comparative study to increase stability and efficiency," Energy and Buildings, art. 115454, 2025, doi:10.1016/j.enbuild.2025.115454; S. Alamouti, "Quantifying Energy and Cost Benefits of Hybrid Edge Cloud: Analysis of Traditional and Agentic Workloads," arXiv:2501.14823, 2025 [preprint].

[45] C. Preist, D. Schien, and P. Shabajee, "Evaluating Sustainable Interaction Design of Digital Services: The Case of YouTube," in Proc. CHI Conference on Human Factors in Computing Systems, 2019, doi:10.1145/3290605.3300627.

[46] Y. Chen et al., "An Overview of Core Coding Tools in the AV1 Video Codec," in Proc. Picture Coding Symposium (PCS), 2018, doi:10.1109/PCS.2018.8456249.

[47] ITU, "New Versatile Video Coding Standard to Enable Next Generation Video Compression," press release, July 2020; InterDigital, "VVC: A Truly Versatile Technology Poised for Growth," 2025.

[48] R. Xu, S. N. Razavi, and R. Zheng, "Edge Video Analytics: A Survey on Applications, Systems and Enabling Techniques," IEEE Communications Surveys & Tutorials, vol. 25, pp. 2951-2982, 2023, doi:10.1109/COMST.2023.3323091.

[49] Y. Wang, C. Yang, S. Lan, L. Zhu, and Y. Zhang, "End-Edge-Cloud Collaborative Computing for Deep Learning: A Comprehensive Survey," IEEE Communications Surveys & Tutorials, vol. 26, pp. 2647-2683, 2024, doi:10.1109/COMST.2024.3393230.

Bhat, M.N. (2026). The Edge as Origin: Five Architectural Transitions Enabling Real-Time Media on Consumer ARM Platforms. South African Computer Journal 38(2), 77–86.

ISSN 1015-7999 (print) ISSN 2313-7835 (online)