WebRTC delivery decision worksheet

Seven questions an architecture call should answer before choosing WebRTC delivery — 2026

Print this page and walk through the questions in order. Each answer narrows the protocol choice; by question seven the architecture is usually clear. Reference article: WebRTC Delivery — From Peer-to-Peer to Scaled Distribution.

1. Latency target

What end-to-end latency does the product actually require? Under 500 ms is the only reason to use WebRTC delivery; 2–3 seconds is LL-HLS territory; 6+ seconds is classic HLS or DASH. Write down the number that lets the viewer's reaction matter.

Answer: _____

2. Peak concurrent viewers

How many viewers will watch the single biggest stream at the same time? Under 5,000 makes WebRTC essentially free; 5K–100K is the hybrid-stack sweet spot; above 100K the broadcast tail must be LL-HLS, MoQ, or DASH.

Answer: _____

3. Device coverage

Which devices must the product reach? Browsers, iOS, and Android all support WebRTC natively. Smart-TV WebRTC support is uneven; Roku has none; tvOS needs a custom app. If smart-TV reach is mandatory, lean on the LL-HLS tier.

Answer: _____

4. DRM requirement

Does the content need studio-grade DRM? WebRTC has no native DRM equivalent. If the content is licensed catalogue or premium VOD, use HLS or DASH with Common Encryption (cbcs). WebRTC delivery suits user-generated, live, and interactive content.

Answer: _____

5. Hybrid-stack tolerance

Is the team prepared to operate a WebRTC tier and an LL-HLS tier in parallel, with a decode-once bridge between them? The hybrid stack is the production reality for any product above ~10K concurrent viewers. If the team cannot run both, pick one tier and accept its trade.

Answer: _____

6. Vendor vs self-host

Will the team use a managed WebRTC delivery service (Cloudflare Stream, Dolby OptiView, LiveKit Cloud, AWS IVS Real-Time, Mux Live) or run its own SFU (mediasoup, Janus, LiveKit Server, Ant Media)? Managed services trade cost for operational simplicity; self-hosting trades operations for control and data residency.

Answer: _____

7. Budget envelope per viewer-hour

What is the all-in budget per viewer-hour? WebRTC delivery costs roughly $0.05–$0.15 per GB egress plus $0.0004–$0.0005 per viewer-minute of SFU compute. LL-HLS costs $0.01–$0.05 per GB and effectively zero compute per viewer. Multiply by audience and duration before committing.

Answer: _____

If three or more answers point to sub-second latency at under 10K concurrent viewers — WebRTC delivery is the right tool. Otherwise, look at LL-HLS or a hybrid stack first.