

Cache Invalidation Playbook

Four tools, four use cases, six pre-flight checks — for any streaming on-call.

Use this playbook when a teammate proposes let's just purge the cache. The four tools below cover ~95% of the work without an invalidation. The four use cases are when a real invalidation is the right tool. The six pre-flight checks are what the on-call should verify before clicking purge.

1. The four tools that handle ~95% of the work

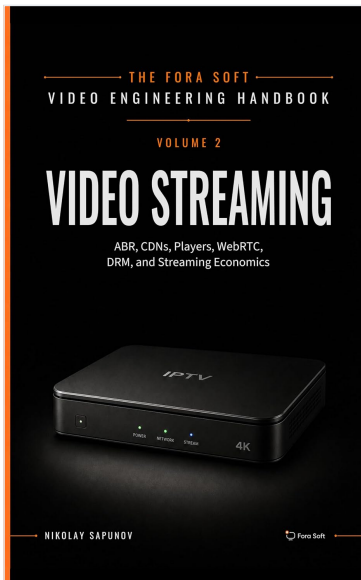
- ☐ **1. Short Cache-Control on manifests**
max-age ≈ half the target segment duration; immutable + max-age=31536000 on segments.
Tool 1 of 4 — designed-in, no operator action.
- ☐ **2. Versioned segment filenames**
When bytes change, rename the path (v1 → v2). The manifest follows; old URL ages out alone.
Tool 2 of 4 — designed-in, no operator action.
- ☐ **3. Segment-timeline rolloff (live)**
Packager drops the segment from the playlist; players stop asking for it within one refresh.
Tool 3 of 4 — designed-in, no operator action.
- ☐ **4. CDN invalidation (paid, rate-limited, last resort)**
CloudFront path: 10–15 min, \$0.005 past 1,000/mo. CloudFront Cache Tag (Apr 2026): < 5 s P95. Fastly: ~150 ms. Akamai: ~5 s. Cloudflare: < 30 s. Google: a few minutes.
Tool 4 of 4 — what is left when tools 1-3 cannot help.

2. The four use cases that actually need invalidation

- ☐ **a. DRM key rotation, mid-flight**
Purge only the EXT-X-KEY URI (HLS) or ContentProtection URL (DASH). Segments untouched.
Triggered by: revoked subscriber entitlement or scheduled rotation firing early.
- ☐ **b. DMCA / GDPR / right-to-be-forgotten takedown**
Remove origin asset first → wildcard invalidation at CDN → leave 410 Gone or 451 at origin.
Triggered by: legal notice; safe-harbour clock ≈ 24 hours.
- ☐ **c. Deployed bug in a single piece of content**
Re-encode under a new versioned path → update manifest → invalidate the manifest only.
Triggered by: bad audio sync, ad-break misfire, ladder error on one title.
- ☐ **d. Live stream that went genuinely wrong**
Stop ingest → let segment-timeline roll off → pull VOD origin → invalidate wildcard → publish clean replacement.
Triggered by: legal pullback, accidental publication, presenter incident.

3. Six pre-flight checks before the on-call clicks purge

- ☐ **1. Origin fixed first**
The bad file at origin is deleted, replaced, or returns 410/451 before any invalidation runs.
Otherwise the very next request pulls the bad bytes back into cache and the invalidation is wasted.
- ☐ **2. Wildcard scope reviewed**
Path includes at most one trailing wildcard; never /*. Prefix must be specific to the affected asset.
A /* invalidation collapses the global hit ratio of the distribution for the next propagation window.
- ☐ **3. Versioning considered first**
Confirmed there is no way to solve this by renaming the affected files and updating the manifest.
Versioning is free and surgical; invalidation is paid and lossy.
- ☐ **4. Cache key matches the URLs being invalidated**
Signed-URL parameters, Authorization headers, Vary fields excluded from the cache key for these paths.
Otherwise the purge can fail to match because the cache key includes things the purge URL does not.
- ☐ **5. CDN propagation time understood**
Documented expectation for the operation: path (10–15 min on CloudFront, ~5 s on Akamai, ~150 ms on Fastly) or tag (< 5 s on CloudFront, < 30 s on Cloudflare).
Communicate the window to support and to anyone driving customer comms.
- ☐ **6. Quota and rate-limit headroom checked**
Below the per-account per-minute cap; CloudFront in-flight invalidations < threshold; budget for paths above the free tier signed off.
Real incidents fail because the quota was already used by a misconfigured deploy job.



GO DEEPER · VOLUME 3 OF 7

Video Streaming: A Complete Guide to Video Delivery

From ABR and CDNs to Players, WebRTC, DRM, and the Economics of Streaming

By Nikolay Sapunov, CEO at Fora Soft · 1,182 pages · Kindle & paperback on Amazon

[Read it on Amazon →](#)
amazon.com/dp/B0HH76ZYWC

This playbook lists the four tools that make purges rare, the four cases that need one and six checks before clicking purge. Volume 3 covers the delivery layer behind it: CDN architecture and caching, shield tiers and cache behaviour, DRM key rotation, player recovery and the economics of egress.

WHAT YOU WILL FIND INSIDE

- CDN architecture and caching
- Player behaviour when segments change
- DRM keys and license delivery
- Egress cost and delivery economics

THE FULL HANDBOOK · 7 VOLUMES · 9,000+ PAGES · ALL ON AMAZON

Vol. 1 · Video Encoding — pixels and color to H.264, HEVC, AV1, quality metrics and AI

Vol. 2 · Video Streaming I — network transport, contribution and ingest, HLS, MPEG-DASH, CMAF

Vol. 3 · Video Streaming II — ABR, CDNs, players, WebRTC, DRM and the economics of delivery

Vol. 4 · Audio for Video I — PCM, codecs, loudness, accessibility, real-time voice processing

Vol. 5 · Audio for Video II — jitter buffers, lip sync, Atmos, quality metrics, neural codecs

Vol. 6 · AI for Video I — models and inference, computer vision, speech, search, generative video

Vol. 7 · AI for Video II — real-time pipelines, agents, serving, cost, compliance, verticals

On-call keeps purging the CDN to fix streaming incidents?

We will review your TTLs, versioning and purge history and design a delivery setup where invalidation is the rare exception. 250+ video products shipped since 2005.

[Book a 30-min call →](#)
[WhatsApp →](#)
[Email us →](#)