

Hybrid ABR Tuning & Pitfalls Sheet

Four knobs, three named algorithms, four failure modes — the working reference for shipping hybrid ABR.

1. The four knobs

λ_s (switch)	Penalty per kbps of quality change between segments. Default 1.0 (MPC paper). Production: 0.5 - 3.0. Higher = smoother, lower average bitrate.
λ_r (rebuffer)	Penalty per predicted stall second. Default 4.3 (MPC paper). Production: 5 - 10. Always $\gg \lambda_s$ — rebuffer is the worst QoE event.
Horizon N	Segments planned ahead. Default 5 (~20 s with 4 s segments). Smaller N for low-latency; larger N only if predictor is very accurate.
Predictor	Throughput forecast. Baseline: harmonic mean, K=20 (Festive). Upgrades: CS2P HMM, Kairos streaming-aware predictor.

2. Deployment defaults

Deployment	λ_s	λ_r	N	Predictor
Premium VoD, fibre	1.0	5.0	5	harmonic, K=20
Mobile-first OTT	1.5	8.0	4	CS2P HMM or harmonic
Live, classic latency	2.0	10.0	3	harmonic, K=10

3. Where hybrid wins

- Premium VoD on residential lines — rebuffer ratio under 0.4% on premium services.
- Predictable networks (home fibre, fixed wireless) — CS2P clustering and HMM thrive.
- Switch-rate-sensitive apps — λ_s cuts visible quality changes 2-5x at same bitrate.
- Per-title / per-shot ladders — clean utility surface for the optimisation.

4. The three named algorithms

Festive (2012)	Harmonic-mean estimator, stability counter, randomised scheduling. Fairness across competing players. Jiang, Sekar, Zhang at CoNEXT.
MPC (2015)	QoE = quality - λ_s -switches - λ_r -rebuffer over horizon N. FastMPC vs RobustMPC. Yin, Jindal, Sekar, Sinopoli at SIGCOMM.
CS2P (2016)	Session clustering + HMM throughput predictor. 40-50% lower forecast error. Pairs with MPC for 14% QoE gain. Sun et al. at SIGCOMM.
Kairos (2025)	MPC-based with a streaming-aware predictor that models buffer-throughput coupling. Current state of the art.

5. Four failure modes & fixes

Bad predictor	Volatile networks (5G, public Wi-Fi, satellite) defeat point forecasts. Fix: RobustMPC interval forecast, CS2P, or Kairos.
Compute cost	M^N candidates per segment painful on low-end devices. Fix: precomputed lookup table indexed by (buffer, throughput, current rung).
Tuning hell	Three weights, horizon, predictor params. Fix: ship published defaults first, A/B at scale before tuning.
Low-latency target	Sub-3-s buffer leaves no horizon room. Fix: switch to L2A or LoL+ for LL-HLS / LL-DASH.

6. Pitfalls to avoid

- Tuning λ_s / λ_r without measuring rebuffer + switches + average bitrate together.
- Trusting a point throughput forecast with no fallback for prediction-error tails.
- Picking horizon N > 5 — long horizons amplify forecast error.
- Disabling Festive's stability counter to chase responsiveness.
- Benchmarking on a stable wired trace — hybrid's gain shows up on real variability.