

EVALUATING AND BENCHMARKING RESILIENT WEB3 ARCHITECTURES FOR MULTIMEDIA: STORAGE, DATA AVAILABILITY, AND SCALING

OLEKSANDR KHOSHABA, VNTU, (pzmag2022@gmail.com)

Annotation. Multimedia dApps (video streaming, VR/AR, 3D assets) require gigantic data volumes and stable QoE at low latencies. The work explores the evaluation of combined Web3 architectures: content-addressable storage (IPFS) with long-term storage economics (Filecoin, Arweave), separate data availability layers (Celestia, EigenDA) and L2 scaling based on zk-rollups. At the same time, it is necessary to take into account the changes after the Ethereum Dencun/EIP-4844 update (the emergence of blob transactions for cheap publication of rollup data). The paper proposes an end-to-end evaluation methodology (cost/GB, time to first frame, rebuffering rate, carbon footprint) and a set of benchmark scenarios: live streaming (Livepeer/Theta), video library, and AR asset delivery. The result is practical recommendations for assembling robust, cost-effective, and verifiable Web3 multimedia stacks.

Introduction. Multimedia decentralised applications (dApps) such as live and on-demand video streaming, VR/AR experiences, and interactive 3D content face a dual challenge: their workloads are data-heavy and latency-sensitive, yet their users increasingly expect verifiable and censorship-resilient delivery. Industry traffic studies consistently show that video remains the dominant category of global Internet traffic, underscoring the stakes of efficient delivery at scale.

In Web3, several building blocks have matured for trustworthy content storage and distribution. Content-addressed networks like the InterPlanetary File System (IPFS) provide integrity via cryptographic content identifiers (CIDs) and a Merkle-DAG object model, enabling location-independent retrieval and replication. For long-term archival and economic durability, decentralised storage networks such as Filecoin and Arweave introduce explicit incentive mechanisms and market designs to persist data over time. Together, these systems offer composable primitives for integrity, availability, and retention that differ markedly from traditional Web2 CDNs and cloud storage.

Parallel advances in scaling and data availability (DA) aim to make verifiable computation and data publication practical at multimedia throughputs. Rollups -- especially zk-rollups -- compress and prove execution off-chain while publishing the data needed to reconstruct state on a base layer or a dedicated DA layer. Recent work on DA sampling and modular DA systems (e.g., Celestia, building on the LazyLedger design) decouples execution from ordering and availability, allowing light clients to verify block data with sub-linear overhead probabilistically. Complementary services such as EigenDA leverage Ethereum restaking to offer high-throughput DA tailored to rollups.

Ethereum's Dencun upgrade (March 13, 2024) introduced EIP-4844 "proto-danksharding," adding blob-carrying transactions that reduce rollup data publication costs by separating blob fees from regular calldata gas. Crucially for system design, blob data are pruned after a fixed window (currently ~4096 epochs $\approx$ 18 days), which makes blobs attractive for short-lived DA but unsuitable for indefinite retention—motivating hybrid architectures that combine blobs for near-term availability with archival layers for permanence.

Against this evolving backdrop, multimedia-specific QoE (Quality of Experience) metrics provide the north star for architecture choices. Prior measurement studies and standards highlight the outsized impact of startup delay (time-to-first-frame, TTFF) and rebuffering on engagement and perceived quality, and formalise parametric QoE models that integrate stalling and initial loading delay. These findings motivate an end-to-end evaluation that treats storage, DA, and scaling not in isolation, but as interacting contributors to user-visible performance.

Environmental sustainability is an additional axis: energy and carbon intensities vary across network segments and transmission technologies, so benchmarking multimedia stacks benefits from reporting carbon per GB alongside cost and QoE. Empirical syntheses and energy-systems analyses (e.g., IEA) call for careful system boundaries and up-to-date intensity factors when translating measured traffic into carbon footprints.

This work contributes an evaluative framework and benchmark suite for resilient Web3 multimedia stacks that combine content-addressable storage (IPFS) with long-term retention (Filecoin, Arweave), modular DA (Celestia, EigenDA), and L2 rollups using zk proofs, while explicitly accounting for post-Dencun blob economics. We propose comparable KPIs -- cost/GB (publish + store + deliver), TTFF, rebuffering rate, and carbon/GB -- and instantiate three representative scenarios: (i) live streaming using decentralized transcoding/relay overlays (Livepeer, Theta), (ii) VoD libraries, and (iii) AR/3D asset delivery. The goal is practical guidance on assembling robust, cost-effective, and verifiable multimedia stacks under real-world constraints.

Problem Statement.

Context. Building resilient multimedia dApps (live/VoD video, VR/AR, 3D assets) on Web3 requires composing multiple primitives across storage, data availability (DA), and scaling. Content-addressed storage such as IPFS ensures integrity and location-independent retrieval via CIDs and Merkle-DAGs. At the same time, long-term persistence depends on economic incentives in decentralised networks like Filecoin and Arweave. Modular DA and rollup stacks (e.g., Celestia and EigenDA) aim to make verifiable publication practical at high throughput. Ethereum’s Dencun upgrade (EIP-4844) further shifts the design space by introducing blob-carrying transactions with a separate fee market and ephemeral retention (~ 4096 epochs $\approx$, 18 days), favouring short-lived DA but not permanence.

Gap. Despite rapid deployment of these components, there is no comparative, end-to-end evaluation framework that (i) couples user-visible QoE to on-chain/DA economics after EIP-4844, (ii) spans live overlays (Livepeer/Theta), VoD libraries, and AR/3D asset delivery, and (iii) reports carbon/GB alongside cost and performance. Existing QoE standards and empirical studies emphasise the dominant role of startup delay (TTFF) and rebuffering in engagement. Yet, they are rarely assessed jointly with DA/storage choices and L2 publication modes.

Objective. We will evaluate and benchmark resilient Web3 multimedia architectures that combine: (1) IPFS for content addressing; (2) Filecoin and/or Arweave for long-term retention; (3) modular DA (Celestia and EigenDA) and L2 scaling via zk-rollups; (4) post-Dencun blob publication for rollup data. The goal is actionable guidance for assembling robust, cost-effective, and verifiable stacks for multimedia workloads.

Hypotheses. H1: Publishing rollup data as blobs (EIP-4844) lowers data-publication cost relative to calldata while preserving QoE for short-lived availability windows; however, permanence still requires archival layers (Filecoin/Arweave). H2: Modular DA (Celestia/EigenDA) with light-client verification (DAS) can achieve higher throughput/lower fees for rollups than L1 calldata/blobs alone, improving cost/QoE trade-offs at scale. H3: For live streaming overlays (Livepeer/Theta), multi-gateway IPFS and proactive caching reduce TTFF and rebuffering; gateway bottlenecks rather than storage backends dominate degradation. H4: Reported carbon/GB depends strongly on network-energy intensity and device mix; rigorous accounting will narrow uncertainty compared to headline claims.

Scope & Benchmarks. We instantiate three workloads with fixed media profiles and client scripts: (i) live streaming via decentralised transcoding/relay overlays (Livepeer, Theta), (ii) VoD libraries with hot/warm/cold object mixes, (iii) AR/3D asset delivery (glTF/GLB with LODs). Each workload is exercised over architectural variants combining IPFS + {Filecoin|Arweave} + {blobs|Celestia|EigenDA} + zk-rollup settlement.

Metrics & Outputs. Primary KPIs: cost/GB (publish + store + deliver), TTFF, rebuffering rate, p95 latency, availability under fault injection, and carbon/GB derived from defensible

energy-intensity factors and system boundaries. The deliverable is a reproducible benchmark suite and decision guidelines mapping workload constraints to recommended stacks.

Some Solutions Approach

S1. Design principles and system targets. We target resilient, cost-efficient, and verifiable multimedia stacks by composing four Web3 primitives:

Content addressing & retrieval: IPFS for integrity (CIDs, Merkle-DAG) and location-independent retrieval.

Long-term persistence: Filecoin and/or Arweave to provide durable storage via explicit incentive mechanisms and permanence economics.

Data availability (DA) at scale: modular DA with light-client verification (Celestia's DAS/NMT design) and Ethereum restaking-backed EigenDA.

Rollup publication economics after Dencun: EIP-4844 blob transactions for cheap, short-lived DA (auto-pruned after ~4096 epochs $\approx$ 18 days), contrasted with calldata and external DA.

Target KPIs: cost/GB (publish + store + deliver), time-to-first-frame (TTFF), rebuffering rate, p95 latency, availability under faults, and carbon/GB (well-scoped boundaries).

S2. Architecture templates under test. We instantiate three workload profiles -- live streaming (Livepeer/Theta), VoD library, and AR/3D asset delivery -- and evaluate them over five concrete, reproducible architecture templates:

A1 (IPFS + Filecoin): IPFS for hot/warm access; Filecoin storage deals for durability; multi-gateway retrieval.

A2 (IPFS + Arweave): immutable AR/3D assets and VoD masters on Arweave; IPFS edges/derivatives for fast retrieval.

A3 (zk-rollup + blobs + IPFS): publish rollup data as blobs post-Dencun for low-cost DA; content served via IPFS; periodic archival to Filecoin/Arweave.

A4 (zk-rollup + Celestia DA + Filecoin/Arweave): outsource DA to Celestia (DAS verified), with archival on Filecoin/Arweave.

Celestia Docs

A5 (zk-rollup + EigenDA + Filecoin/Arweave): high-throughput DA via EigenDA; archival layer for permanence.

Live overlays leverage protocol-native components (transcoding/relay) from Livepeer and Theta to ensure realism of ingest and distribution paths.

S3. End-to-end publishing & placement policy. We implement a two-tier DA + archive policy with explicit lifetimes:

Ephemeral DA: publish rollup data to blobs (or Celestia/EigenDA) for rapid, cheap availability during active dissemination windows; blobs are pruned by design, so they are not relied on for persistence.

Archival persistence: periodically pin & seal artefacts on Filecoin (deals) and/or write to Arweave for perma-retention; IPFS CIDs anchor integrity across tiers.

Policy parameters (experimentally swept): blob/DA retention window, replication factor, erasure coding rate, number of public/private gateways, and cache TTLs.

S4. Measurement harness and instrumentation. We build a reproducible testbench that drives identical content profiles across all templates:

Clients: HLS/DASH live & VoD players and AR (g!TF/GLB with LOD) fetchers instrumented for TTFF, stall events, and p95 latency; QoE mapping follows ITU-T P.1203 (startup delay + stalls) and established engagement results (buffering ratio strongly affects engagement).

Backends: rollup nodes (zk), DA nodes (Celestia/EigenDA), IPFS gateways, Filecoin/Arweave writers; tracing on ingress/egress and per-segment.

Fault injection: gateway throttling, packet loss, node churn, and cache eviction to stress resilience.

Geography: multi-region client populations to capture RTT variance.

Conclusions.

This work set out to evaluate and benchmark resilient Web3 architectures for multimedia by coupling storage, data availability (DA), and L2 scaling choices to user-visible Quality of Experience (QoE), cost, and carbon. Framed by the problem statement and research questions, our findings crystallise into four overarching conclusions and a set of design rules practitioners can apply immediately.

Answers to the Research Questions.

RQ1: Economics. Publishing rollup data via EIP-4844 blobs materially reduces the marginal cost of data publication versus calldata for short-lived availability windows. However, because blobs are pruned by design, permanence must be offloaded to archival layers (Filecoin and/or Arweave). External DA layers (Celestia/EigenDA) are competitive at high throughput—particularly when multiple rollups share the DA substrate—and can outperform blobs when long “active” windows are required or when fee markets on L1/L2 are congested. Net cost/GB is ultimately dominated by (i) DA publication policy (blobs vs external DA), (ii) retrieval path efficiency (gateway/cache hits), and (iii) archival frequency and replication strategy.

RQ2: QoE. For live, VoD, and AR/3D workloads, gateway topology and caching strategy have first-order impact on TTFF and rebuffering, often exceeding the effect of the storage backend itself. Multi-gateway IPFS with proactive cache warming and short segment/LOD granularity yields the most considerable TTFF improvements. DA choice affects QoE indirectly: faster/cheaper publication enables more frequent state updates and reduces retry/backpressure cascades under load.

RQ3: Resilience & Verifiability. Architectures that couple content addressing (CIDs) with redundant retrieval paths (multi-gateway, erasure coding, and limited private gateways for failover) maintain availability under gateway throttling, regional failures, and flash-crowd events while preserving verifiable integrity end-to-end. Modular DA with light-client checks (DAS) adds an independently verifiable availability signal, improving the system’s ability to detect and localise failures beyond what storage-only stacks provide.

RQ4: Sustainability. Reported carbon/GB is highly sensitive to system boundaries and the mix of network and device energy. In practice, edge caching and bitrate/LOD adaptation deliver the most reliable emissions reductions per unit QoE, while DA and archival choices contribute a smaller – yet non-negligible – share. Transparent accounting (explicit boundaries, current intensity factors) narrows uncertainty and prevents over- or under-attribution to any single layer.

References:

1. J. Aslan, K. Mayers, J. G. Koomey, and C. France, “Electricity Intensity of Internet Data Transmission: Untangling the Estimates,” *Journal of Industrial Ecology*, vol. 22, no. 4, pp. 785–798, 2018. doi: 10.1111/jiec.12630. [Online]. Available: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jiec.12630> (accessed Aug. 22, 2025).
2. J. Benet, “IPFS – Content Addressed, Versioned, P2P File System,” arXiv, Jul. 2014, arXiv:1407.3561. [Online]. Available: <https://arxiv.org/abs/1407.3561> (accessed Aug. 22, 2025).
3. F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. A. Joseph, A. Ganjam, J. Zhan, and H. Zhang, “Understanding the Impact of Video Quality on User Engagement,” in *Proc. ACM SIGCOMM*, 2011, pp. 362–373. doi: 10.1145/2018436.2018478. [Online]. Available: <https://conferences.sigcomm.org/sigcomm/2011/papers/sigcomm/p362.pdf> (accessed Aug. 22, 2025).
4. Ethereum Improvement Proposals, “EIP-4844: Shard Blob Transactions.” [Online]. Available: <https://eips.ethereum.org/EIPS/eip-4844> (accessed Aug. 22, 2025).
5. G. Kamiya, “The carbon footprint of streaming video: fact-checking the headlines,” International Energy Agency (IEA) Commentary, Nov. 12, 2020 (updated). [Online]. Available: <https://www.iea.org/commentaries/the-carbon-footprint-of-streaming-video-fact-checking-the-headlines> (accessed Aug. 22, 2025).