Technology White Paper

MEDIA OVER QUIC

The Next Generation Protocol for Real-Time Video Streaming

A comprehensive guide for CTOs and technical leaders on why Media over QUIC is revolutionizing real-time video delivery, overcoming traditional streaming limitations, and enabling ultra-low latency experiences at global scale.

created and brought to you by

nano COSMOS

www.nanocosmos.net

Real-Time Streaming Technology | Protocol Innovation | Enterprise Implementation

REAL-TIME VIDEO THAT SIMPLY WORKS

Table of Contents

01 Executive Summary

The Real-Time Video Revolution

Traditional live streaming is non-interactive for lean-back scenarios. The audio/video content does not need to be at the viewer in real-time. Delays between the source and destination are at least 6 seconds, typically even 30 seconds (also known as latency).

Almost all implementations are based the streaming protocols HLS and DASH, built on TCP foundations. HLS and DASH have economic and technical benefits, as CDNs can easily forward and cache chunks of video data in their network and make distribution to large audiences efficient.

However, there is a growing demand for applications with audience interaction – use cases like live betting, interactive auctions, virtual town halls, and collaborative applications. They all rely on real-time interaction where every fraction of a second matters. Ideally, end-to-end latency between the presenter and the audience stays below one second (sub-second) or lower.

The demand for interactive real-time applications has been fulfilled by vendors like nanocosmos who have managed to provide ultra low latency, real-time streaming platforms for interactive use cases. These implementations are built on different standards than HLS and DASH, with transport protocols like Websockets or HTTP-based ultra-low-latency live streaming.

Browser-based applications for playback on any device need to use what the browser can provide and rely on what is available in the browser technology stack. This has been either HTTP, Websockets, or WebRTC, and the emerging QUIC and Webtransport for many browsers.

Media Over QUIC (MoQ) represents a fundamental shift in how we approach real-time video delivery, offering more reliable sub-second latency while maintaining broadcast-quality video and global scalability. In browser-based applications, the Webtransport protocol enables a flexible API for media transport over QUIC.

200-300ms

Ultra-Low Latency

End-to-end latency achievable with MoQ vs 6-30 seconds with traditional protocols

99.9%

Network Resilience

Reliability in challenging network conditions with intelligent packet handling

∞

Global Scale

One-to-many distribution without WebRTC's peer-to-peer limitations

Important: MoQ is Not Plug-and-Play

While MoQ offers revolutionary capabilities, it requires sophisticated implementation, fallback mechanisms, and deep integration expertise. This white paper explores both the opportunities and challenges of adopting MoQ technology.

02 The Real-Time Video Challenge

Why Traditional Protocols Fall Short

The streaming landscape was built for a different era. When Apple pioneered HLS in 2009 and DASH followed shortly, the primary goal was reliable delivery live content just for watching, without any interaction. Viewers accepted buffering time and delays in exchange for consistent high playback quality.

Many applications today demand real-time interaction. Live sports betting requires sub-second latency to maintain fair odds. Interactive auctions need immediate bid updates. Corporate webinars require instant Q&A responses. Traditional protocols simply weren't designed for this reality.

Latency Bottlenecks

Latency Bottlenecks Traditional HLS/DASH protocols typically use 2+ second segments, creating minimum 6-second delays, which makes real-time interaction impossible.

Business Impact:

Business Impact: Lost revenue in live betting, poor user engagement in auctions, frustrated users in interactive sessions.

Network Instability

Network Instability TCP's 'polite' protocol waits for acknowledgments, creating buffering and quality drops when networks are unstable. The TCP implementation of waiting and resending is rather outdated and does not work well in unreliable networks which are typical in mobile and remote areas.

Business Impact:

Business Impact: Frequent rebuffering on mobile networks, degraded quality during peak usage, poor experience for remote users.

Scalability Limitations

WebRTC excels at peer-to-peer but struggles with large-scale broadcasting due to bandwidth and infrastructure requirements, as well as bitrate adaption.

Business Impact:

Limited audience size, high infrastructure costs, complex fallback requirements for different devices.

Quality vs Speed Trade-offs

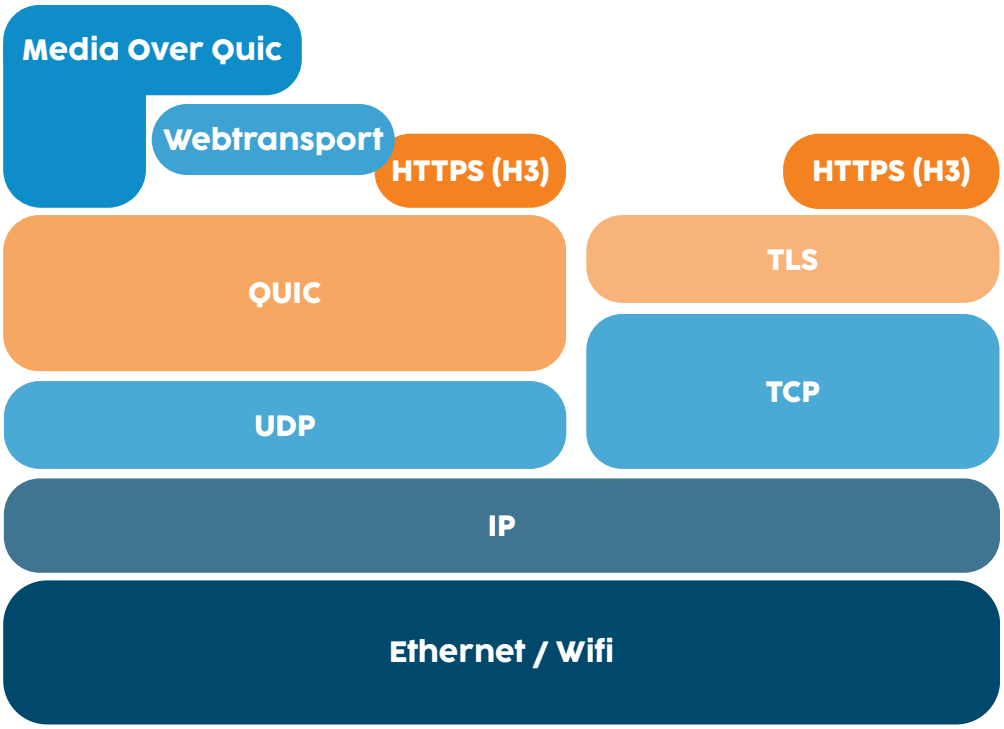
Current solutions force choosing between broadcast quality or low latency—you can't have both reliably.

Business Impact:

Compromised video quality for interactive applications or unacceptable delays for high-quality streams.

03 Understanding Media Over QUIC

A Protocol Built for Real-Time Media



Media Over QUIC (MoQ) represents a fundamental reimagining of how real-time media should be transported over the internet. Unlike traditional protocols that were adapted for video, MoQ was designed from the ground up specifically for real-time media delivery.

Built on QUIC (Quick UDP Internet Connection), MoQ combines the speed of UDP with the reliability features traditionally associated with TCP, creating a transport layer optimized for the demands of modern streaming applications.

Ultra-Low Latency

Built on QUIC's UDP foundation with intelligent error correction, eliminating TCP's acknowledgment delays.

Built-in Encryption

All QUIC connections are encrypted by default, providing enterprise-grade security without additional overhead.

Stream Multiplexing

Multiple independent streams prevent head-of-line blocking, ensuring smooth delivery even with packet loss.

Adaptive Prioritization

Publishers can define priority levels for different tracks, ensuring critical data delivers first during congestion.

How MoQ Solves Real-World Problems

Network Resilience

When users switch from WiFi to 5G to 3G, MoQ's connection management maintains stream continuity without rebuffering, unlike TCP-based protocols that must re-establish connections.

Intelligent Quality Management

Publishers can prioritize audio over video during congestion, or ensure critical data streams (like auction bids) always take precedence over less critical content.

Flexible Latency Control

Unlike WebRTC's "latency-first" approach that sacrifices quality, MoQ allows granular control—viewers can choose their position relative to the live edge based on their specific needs.

04 MoQ vs Traditional Protocols

Comprehensive Protocol Analysis

Understanding where MoQ fits in the streaming protocol landscape requires examining the fundamental trade-offs each technology makes. This analysis compares MoQ against the dominant protocols in today's streaming ecosystem.

MoQ vs. HLS vs. Websocket

MoQ is very different to HLS: it provides a continuous live stream, compared to segments of video chunks for HLS. Also, MoQ is using QUIC for transport compared to TCP with HLS. This all means: lower latency, better network resilience.
MoQ with Webtransport is closer to Websocket. Webtransport is using UDP, Websockets TCP.

MoQ vs. WebRTC

WebRTC was introduced 2013 by Google as part of their Chrome browser. It took many years for a standard which was introduced 2021. WebRTC was designed as a peer-to-peer communication protocol for telephony, not for live streaming. It is a monolithic API which is very complex under the hood with little control of the video compression and streaming formats. Also connection handshake is difficult to handle with a lot of STUN and TURN server requirements. Quality is limited due to little control, it always favors realtime communication over quality and possible drops frames during a session. Also bitrate adaption is difficult to handle. For live streaming to large audiences, a complex server infrastructure is required.
MoQ was designed to work in large scale one-to-many situations and is more flexible in video and stream configuration.

Detailed Protocol Comparison

Feature	HLS/ DASH	WebRTC	MoQ	MoQ Advantage
Typical Latency	6-30 seconds	100-500ms	300ms	Consistent sub-second with quality
Video Quality	Excellent	Limited (no B-frames)	Broadcast -grade	High quality + low latency
Scalability	Global CDN	P2P limitations	One-to-many, unlimited	CDN-friendly architecture
Network Handling	Buffering on issues	Frame drops	Intelligent recovery	Maintains stream continuity
Implementation	Simple	Complex (STUN/TURN)	Requires expertise	Modular, CDN-compatible

Why Traditional Protocols Struggle

- HLS requires typically 2+ second segments, creating unavoidable latency of min 6 seconds
- TCP's head-of-line blocking causes buffering cascades
- WebRTC's complexity makes large-scale deployment challenging
- Quality vs latency trade-offs limit real-time applications

MoQ's Strategic Advantages

- Sub-second latency with broadcast-quality video
- Independent streams prevent blocking issues
- CDN-compatible architecture for global scale
- Intelligent priority management during congestion
- Simpler, more flexible browser implementation with Webtransport API

Technical Deep Dive: QUIC vs TCP

TCP Limitations

- Single stream support
- Multi-step handshake increases connection time
- Single stream vulnerable to head-of-line blocking
- Full retransmission on packet loss
- Optional encryption adds overhead

QUIC Advantages

- Multi stream support
- Single-step handshake with built-in encryption
- Independent streams prevent blocking
- Selective retransmission of lost packets only
- Connection migration for mobile scenarios

Browser implementation

There is a web standard for data transport over QUIC and HTTP/3 called Webtransport.
This API is available in all modern browsers (except Safari yet), making it possible to take advantage of QUIC as a media transport.

Conceptually, WebTransport can be compared to WebSocket but utilizes QUIC instead of TCP, providing benefits such as stream multiplexing and support for datagrams—features that enhance performance and efficiency for real-time communication. Despite these conceptual similarities, WebTransport and WebSocket differ significantly in their underlying protocols.

MoQ Standardization

There is an ongoing standardization process by IETF to define MoQ as a transport for media and other data over QUIC. It is designed as a publish/subscribe protocol for great flexibility for more than just media (audio/video), but also for other types of real-time data, like for internet of things IoT.

The standardization process is ongoing as of end 2025 and will lead to the MoQ-T (Moq-Transport) standard. There is discussion to get rid of the term “Media over QUIC” and just name it MoQ or “Many things over QUIC”. This standard will be able to cover a lot of use cases, with many details unclear.

nanocosmos decided not to wait for this standardization process but instead make “media over QUIC” available as quickly as possible for the industry to validate it’s promises and show that it is a reliable and useful format for enabling and improving real-time streaming use cases.

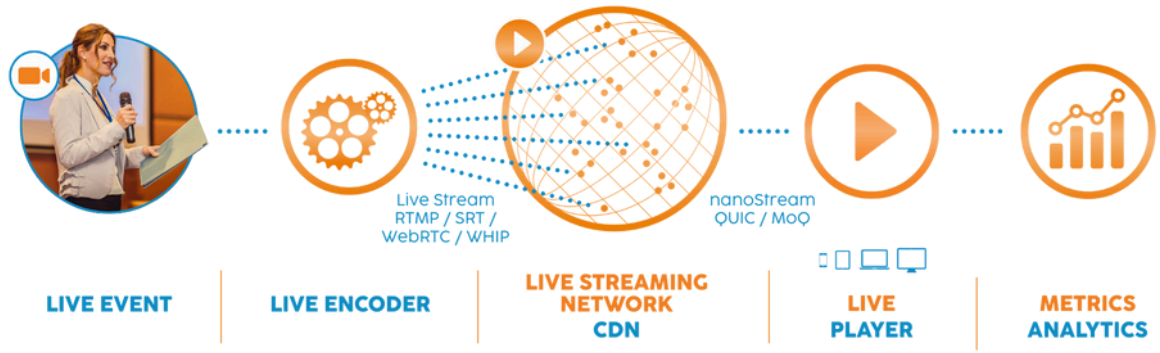
The nanocosmos philosophy:

- We don't wait for standards to be finalized
- We implement and solve industry needs today
- Proven approach 25+ years of early adoption success
- Our customers need solutions now
- We are open to connect and enable valuable use cases

05 Technical Architecture & Workflow

End-to-End MoQ Streaming Architecture

Implementing MoQ successfully requires understanding the complete streaming workflow from ingest to playout. Each stage of the pipeline plays a critical role in achieving the ultra-low latency and high quality that MoQ promises.



STEP 1

Ingest

Content capture and encoding

Live cameras, encoders, or browser-based capture send streams to the platform

WebRTC (WHIP) · RTMP · SRT

STEP 2

Distribution

CDN and edge processing

Global CDN with origin-edge architecture using MoQ for ultra-low latency middle-mile transport

Media over QUIC · WebTransport · HTTP/3

STEP 3

Playout

End-user delivery

Intelligent player auto-selects optimal protocol based on browser capabilities and network conditions

Fallback to TCP based protocols, HLS and/or Websocket · WebTransport · MoQ

Protocol Selection at Each Stage

Ingest Stage

Primary:
RTMP/SRT for professional equipment, flexible selection of video profiles

Browser:
WebRTC (WHIP) for web-based publishing

Mature, stable protocols ensure reliable content capture from various sources

Distribution Stage

Primary:
Media over QUIC (MoQ) / WebTransport

Fallback:
HTTP, Websocket

MoQ shines in CDN middle-mile transport, reducing latency between origin and edge

Playout Stage

Modern:
WebTransport + MoQ

Fallback:
TCP-based protocols WebSocket or HLS

Intelligent player selects optimal protocol based on browser support and network conditions

Critical Implementation Considerations

Browser Compatibility

- Chrome/Edge: Full WebTransport support
- Firefox: Support in development
- Safari: Limited experimental support

Infrastructure Requirements

- UDP traffic must be allowed (firewalls)
- Fallback mechanisms essential
- Real-time analytics integration

Real-time interaction / Audience Engagement:
Q&A, Polls, Voting, Surveys, Feedback, Comments, Bidding, Betting

06 Industry Use Cases

Where MoQ Makes the Difference

MoQ's impact varies significantly across different industries and applications. Understanding where ultra-low latency provides the greatest business value helps organizations prioritize their streaming technology investments.

Revenue Critical
Applications where latency directly impacts revenue generation

Engagement Critical
User experience significantly improved by reduced latency

Instant Response
Professional requirements for reliable, low-latency delivery

Live Sports Betting & iGaming

< 500ms | Critical Impact

Real-time odds updates and bet placement where latency directly affects fairness and revenue.

Key Challenges:
- Synchronized odds across millions of users
- Fraud prevention with instant updates
- Regulatory compliance for fair play

MoQ Benefits:
✓ Sub-second bet confirmation
✓ Synchronized live odds
✓ Global scale without WebRTC complexity

Live Commerce & Auctions

< 1 second | High Impact

Interactive shopping experiences where real-time bidding and instant responses drive engagement.

Key Challenges:
- Bid synchronization across viewers
- Inventory updates in real-time
- Payment processing integration

MoQ Benefits:
✓ Instant bid visibility
✓ Real-time inventory updates
✓ Enhanced buyer confidence

Interactive Education & Webinars

< 2 seconds | Medium Impact

Real-time Q&A, polling, and collaborative learning experiences requiring immediate interaction.

Key Challenges:
- Audience participation at scale
- Content synchronization
- Multi-way communication

MoQ Benefits:
✓ Instant Q&A responses
✓ Real-time polling results
✓ Enhanced engagement

Public Safety & Surveillance

< 300ms | Critical Impact

Emergency response and security monitoring where delays can have serious consequences.

Key Challenges:
- Mission-critical reliability
- Multi-source video feeds
- Command center coordination

MoQ Benefits:
✓ Real-time situational awareness
✓ Instant alert distribution
✓ Coordinated response capability

Virtual Town Halls & Events

< 1 second | Medium Impact

Large-scale interactive events requiring audience participation and real-time feedback.

Key Challenges:
- Massive audience scale
- Interactive features
- Content delivery reliability

MoQ Benefits:
✓ Live audience interaction
✓ Scalable participation
✓ Professional broadcast quality

Gaming & Esports

< 200ms | High Impact

Competitive gaming streams where viewer experience and fairness depend on minimal latency.

Key Challenges:
- Ultra-low latency requirements
- High-quality video
- Global tournament streaming

MoQ Benefits:
✓ Tournament-grade latency
✓ Broadcast quality
✓ Global audience reach

ROI Considerations by Industry

High ROI Scenarios

$ Live betting platforms with millions in daily volume

$ High-value auction platforms (art, real estate, luxury goods)

$ Mission-critical public safety applications

$ Premium educational platforms with high engagement requirements

07 Implementation Considerations

MoQ is Not Plug-and-Play Technology

Unlike traditional streaming solutions that can be implemented with simple SDK integrations, MoQ requires sophisticated understanding of real-time protocols, browser compatibility matrices, and complex fallback mechanisms.

Organizations considering MoQ must prepare for significant engineering investment and ongoing protocol expertise maintenance.

Browser Support

WebTransport support varies across browsers, requiring robust fallback mechanisms

Mitigation Strategies:

- Maintain WebSocket/HTTP fallbacks
- Monitor browser adoption metrics

Network Infrastructure

Some corporate firewalls still block UDP traffic, affecting QUIC deployment

Mitigation Strategies:

- Provide TCP fallback via WebSockets/HTTP
- Work with enterprise IT teams
- Monitor connection success rates

Development Complexity

MoQ requires deep streaming expertise and cannot be implemented as a simple library replacement

Mitigation Strategies:

- Partner with experienced vendors
- Invest in team training
- Start with pilot projects

CPU Overhead

QUIC processing is currently more CPU-intensive than TCP

Mitigation Strategies:

- Plan for higher server costs
- Monitor hardware acceleration
- Implement efficient caching

MoQ Readiness Assessment

Before implementing MoQ, organizations should honestly assess their readiness across technical, infrastructure, and business dimensions.

Team Expertise

Critical

- Do you have streaming protocol specialists?
- Experience with real-time systems?
- Ability to handle complex fallback logic?

Infrastructure Maturity

High

- HTTP/3 server support available?
- CDN supports QUIC?
- Monitoring systems in place?

Business Requirements

Medium

- Sub-second latency essential?
- Interactive features needed?
- Global audience reach required?

Architecting bespoke MoQ video infrastructure is an expensive endeavor — and one that takes focus away from your content and viewers. The same can be said for combining multiple components, as many hidden costs and complexities come into play.

By going with an end-to-end video platform like nanocosmos, you can offset development costs and differentiate your service faster.

Hidden efforts

"We need a reliable real-time live streaming service"

Comprehensive approach:
"Zero Configuration"

- Automatic MoQ, ULL-HLS, Websocket
- Complex fallback mechanisms
- Browser compatibility issues
- Video / streaming expertise
- CDN integration for 24/7 operation
- Monitoring and analytics

08 nanocosmos MoQ Implementation

"Real-Time Video That Simply Works"

Since 1998, nanocosmos has been working in the audio and video industry, being a pioneer in real-time video solutions and working across several industries reliably for the last ten years. Their comprehensive nanoStream platform integrates MoQ technology seamlessly, providing enterprise-grade streaming without the complexity typically associated with cutting-edge protocols.

Our Comprehensive Approach to MoQ

End-to-End Integration

Unlike vendors offering MoQ as an isolated component, nanocosmos integrates MoQ throughout our entire nanoStream platform—from CDN-level origin-edge distribution to last-mile delivery to playback devices.

- Automatic protocol selection and fallback
- Seamless browser compatibility handling
- Zero-disruption updates to existing workflows

25+ Years of Real-Time Video and Streaming Expertise

Our deep understanding of real-time video challenges ensures MoQ implementation that addresses real-world deployment scenarios, not just laboratory conditions.

- Battle-tested fallback mechanisms
- Enterprise-grade reliability and support
- Continuous optimization based on real usage

MoQT IETF standardization process (still evolving

MoQ Media Over Quic implementations

Global CDN Infrastructure

Worldwide edge servers with 100% uptime commitment and intelligent routing for optimal MoQ delivery

Analytics

Comprehensive QoE monitoring with instant insights on latency, quality, and user engagement metrics

Enterprise Security

End-to-end encryption, content protection, and compliance features for business-critical applications

Seamless Integration

Simple APIs that abstract MoQ complexity while providing full protocol benefits

Making Innovation Accessible

For Business Leaders

- Focus on business applications, not streaming protocol complexity
- Proven enterprise deployment with 24/7 support
- Predictable costs and transparent pricing models

For Development Teams

- Simple player updates enable MoQ benefits without code rewrites
- Comprehensive SDKs and documentation for rapid integration
- Expert technical support from streaming protocol specialists

09 Getting Started with MoQ

Two Paths to MoQ Implementation

Build In-House

Develop MoQ capabilities internally with your engineering team.

✗ 12-18 month development timeline

✗ Requires specialized protocol expertise

✗ Complex browser compatibility management

✗ Ongoing maintenance and optimization burden

Partner with nanocosmos

Leverage our proven MoQ platform and 25+ years of real-time video and streaming expertise.

✓ Immediate deployment with simple integration

✓ Enterprise-grade reliability and support

✓ Automatic fallbacks and compatibility

✓ Continuous optimization and updates

Why Partner with nanocosmos?

While building MoQ capabilities in-house is possible, partnering with an experienced platform provider offers significant advantages for most organizations.

Our clients focus on their core business while we handle the streaming complexity.

Immediate access to proven MoQ implementation Expert technical support and consultation Comprehensive fallback and compatibility handling 24/7 monitoring and enterprise support Rapid deployment without in-house expertise Continuous optimization and protocol updates

Ready to Experience the Future of Streaming?

Experience the innovation — leverage MoQ today!
Test our MoQ demo right in your browser for a quick preview, or unlock the full power of our real-time video platform with a free signup.
Seeing is believing. Discover how real-time streaming can transform your business.

See MoQ technology in action.
Our live demo brings the concepts from this white paper to life — explore the technology firsthand.

Try MoQ Demo Now | **Talk to an Expert**

Questions? Contact our team at innovation@nanocosmos.net

nano COSMOS

REAL TIME VIDEO

THAT SIMPLY WORKS